Encyclopedia of
ACOUSTICS

VOLUME THREE

Architectural Acoustics
Acoustical Signal Processing
Physiological Acoustics
Psychological Acoustics

Edited by
Malcolm J. Crocker

ENCYCLOPEDIA OF ACOUSTICS

EDITORIAL BOARD

Malcolm J. Crocker, *Editor-in-Chief*

Robert E. Apfel, Yale University, New Haven, CT
David T. Blackstock, University of Texas, Austin, TX
Leo L. Beranek, Cambridge, MA
Per V. Brüel, Brüel Acoustics, Denmark
William J. Cavanaugh, Cavanaugh Tocci Associates, Sudbury, MA
Floyd Dunn, University of Illinois, Urbana, IL
Ira Dyer, M.I.T. Cambridge, MA
Tony F. W. Embleton, National Research Council, Ottawa, Canada
Frank J. Fahy, I.S.V.R., Southampton University, England
James L. Flanagan, Rutgers University, Piscataway, NJ
Henning von Gierke, Wright Patterson AFB, OH
David M. Green, University of Florida, Gainesville, FL
Manfred A. Heckl,* Institute of Technical Acoustics, Berlin, Germany
Joseph E. Hind, University of Wisconsin, Madison, WI
William W. Lang, Poughkeepsie, NY
Sir James Lighthill, University College London, England
Bruce Lindsay,* Brown University, Providence, RI
Leonid M. Lyamshev, Academy of Sciences, Moscow, Russia
Richard H. Lyon, M.I.T., Cambridge, MA
Maa Dah-You, Institute of Acoustics, Beijing, P.R.C.
Philip L. Marston, Washington State University, Pullman, WA
Dennis McFadden, University of Texas, Austin, TX
Harry B. Miller, Naval Underwater Systems Center, New London, CT
Alan Powell, University of Houston, TX
Thomas D. Rossing, Northern Illinois University, DeKalb, IL
Andrew F. Seybert, University of Kentucky, Lexington, KY
Theodore J. Schultz,* Schultz & Associates, Boston, MA
W. Dixon Ward, University of Minnesota, MN

*Deceased

ENCYCLOPEDIA OF ACOUSTICS

Volume Three

MALCOLM J. CROCKER, *Editor-in-Chief*
Auburn University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto

This text is printed on acid-free paper.

Copyright © 1997 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If legal advice or other expert assistance is required, the services of a competent professional person should be sought.

Library of Congress Cataloging in Publication Data:

Encyclopedia of acoustics / [edited by] Malcolm J. Crocker
p. cm.
Includes index.
ISBN: 0-471-18006-8 (Volume Three)
ISBN: 0-471-80465-7 (set)
1. Acoustics—Encyclopedias. I. Crocker, Malcolm J.
QC221.5.E53 1997
534—dc21 96-37424

Printed in the United States of America

10 9 8 7 6 5 4 3 2

CONTENTS

Volume One

FOREWORD xv

PREFACE xvii

CONTRIBUTORS xix

PART I GENERAL LINEAR ACOUSTICS 1

1 **Introduction** 3
Malcolm J. Crocker

2 **Mathematical Theory of Wave Propagation** 21
Allan D. Pierce

3 **Ray Acoustics for Fluids** 39
D. E. Weston

4 **Ray Acoustics for Structures** 47
Leopold B. Felsen

5 **Interference and Steady-State Scattering of Sound Waves** 55
Herbert Überall

6 **Speed of Sound in Fluids** 69
Allan J. Zuckerwar

7 **Standing Waves** 81
U. S. Shirahatti and Malcolm J. Crocker

8 **Waveguides** 91
P. O. A. L. Davies

9 **Steady-State Radiation from Sources** 107
A. P. Dowling

10 **Transient Radiation** 127
Peter Stepanishen

11 **Acoustic Interaction between Structures and Fluids** 135
F. J. Fahy

12 **Boundary Waves** 153
Ivan Tolstoy

13 **Acoustic Lumped Elements from First Principles** 161
Robert E. Apfel

14 **Acoustic Modeling: Finite Element Method** 165
A. Craggs

15 **Acoustic Modeling: Boundary Element Methods** 173
A. F. Seybert and T. W. Wu

16 **Acoustic Modeling (Ducted-Source Systems)** 185
M. G. Prasad and Malcolm J. Crocker

PART II NONLINEAR ACOUSTICS AND CAVITATION 191

17 **Nonlinear Acoustics and Cavitation** 193
David T. Blackstock and Malcolm J. Crocker

18 **Some Model Equations of Nonlinear Acoustics** 197
David T. Blackstock

19 **Propagation of Finite-Amplitude Waves in Fluids** 203
D. G. Crighton

20 **Parameters of Nonlinearity of Acoustic Media** 219
E. Carr Everbach

21 **Finite-Amplitude Waves in Solids** 227
Mack A. Breazeale

22 **Nonlinear Standing Waves in Cavities** 237
Alan B. Coppens and Anthony A. Atchley

23 **Nonlinear Effects in Sound Beams** 249
Mark F. Hamilton

24 **Nonlinear Lumped Elements** 257
William E. Zorumski

25 **Cavitation** 263
Werner Lauterborn

26 **Sonochemistry and Sonoluminescence** 271
Kenneth S. Suslick and Lawrence A. Crum

PART III AEROACOUSTICS AND ATMOSPHERIC SOUND 283

27 **Introduction** 285
James Lighthill

28 **Aerodynamic and Jet Noise** 301
Alan Powell

29 **Interaction of Fluid Motion and Sound** 313
J. E. Ffowcs Williams

30 **Acoustic Streaming** 321
N. Riley

31 **Shock Waves, Blast Waves, and Sonic Booms** 329
Richard Raspet

32 **Atmospheric Sound Propagation** 341
Louis C. Sutherland and Gilles A. Daigle

33 **Infrasound** 367
Thomas B. Gabrielson

PART IV UNDERWATER SOUND 373

34 **Introduction** 375
Ira Dyer

35 **Essential Oceanography** 381
F. H. Fisher and P. F. Worcester

36 **Propagation of Sound in the Ocean** 391
William A. Kuperman

37 **Propagation in Marine Sediments** 409
LeRoy M. Dorman

38 **Attenuation by Forward Scattering: Measurements and Modeling** 417
Robert H. Mellen, Hans G. Schneider, and David G. Browning

39 **Volume Scattering in Underwater Acoustic Propagation** 425
Shimshon Frankenthal

40 **Backscattering from Rough Surfaces and Inhomogeneous Volumes** 441
R. P. Chapman, O. Z. Bluy, and P. C. Hines

41 **Sound Radiation from Marine Structures** 459
David Feit

42 **Transient and Steady-State Scattering and Diffraction from Underwater Targets** 469
Louis R. Dragonette and Charles F. Gaumond

43 **Quantitative Ray Methods for Scattering** 483
Philip L. Marston

44 **Target Strength of Fish** 493
Kenneth G. Foote

45 **Fundamental Underwater Noise Sources** 501
William Blake

46 **Ship and Platform Noise, Propeller Noise** 521
Robert D. Collier

47 **Underwater Explosive Sound Sources** 539
Valery K. Kedrinskii

48 Ocean Ambient Noise 549
Ira Dyer

49 Sonar Systems 559
James E. Barger

50 Oceanographic and Navigational Instruments 581
Robert C. Spindel

51 Acoustic Telemetry 591
Josko A. Catipovic

52 Transducers 597
Joseph E. Blue and Arnie Lee Van Buren

53 Nonlinear Sources and Receivers 607
Mark B. Moffett and William L. Konrad

Volume Two

PART V ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND 619

54 Introduction 621
Philip L. Marston

55 Ultrasonic Velocity 629
John H. Cantrell and William T. Yost

56 Ultrasonic Relaxation Processes 641
Henry E. Bass and F. Douglas Shields

57 Phonons in Crystals, Quasicrystals, and Anderson Localization 651
Julian D. Maynard

58 Surface Waves in Solids and Ultrasonic Properties 661
Moisés Levy and Susan C. Schneider

59 Wave Modes in Liquid Helium 673
Gary A. Williams

60 Ultrasonic Instruments for Nondestructive Testing 683
Emmanuel P. Papadakis

61 Thermoacoustic Engines 695
Gregory W. Swift

PART IV MECHANICAL VIBRATIONS AND SHOCK 703

62 Introduction 705
Manfred A. Heckl

63 Experimental and Theoretical Studies of Vibrating Systems 715
Paul J. Remington

64 Vibrations of One- and Two-Dimensional Continuous Systems 735
Manfred A. Heckl

65 **Nonlinear Vibration** 753
Lawrence N. Virgin and Earl H. Dowell

66 **Random Vibration** 765
David E. Newland

67 **Shock Analysis and Design** 785
Thomas L. Geers

68 **Acoustic Emission** 797
Kanji Ono

69 **Seismic Exploration: Acoustic Imaging for Geological Information** 811
A. J. Berkhout

70 **Effects of High-Intensity Sound on Structures** 831
B. L. Clarkson

71 **Vibration Isolation and Damping** 843
Eric E. Ungar

72 **Vibration Measurements and Instrumentation** 857
Michael Möser

73 **Machinery Condition Monitoring** 869
Govindappa Krishnappa

74 **Structure-Borne Energy Flow** 881
G. Pavic

75 **Active Vibration Control** 893
C. R. Fuller

PART VII STATISTICAL METHODS IN ACOUSTICS 909

76 **Introduction** 911
Richard H. Lyon

77 **Response Statistics of Rooms** 913
Mikio Tohyama

78 **Statistical Modeling of Vibrating Systems** 925
Jerome E. Manning

PART VIII NOISE: ITS EFFECTS AND CONTROL 937

79 **Noise Control** 939
Malcolm J. Crocker

80 **Rating Measures, Descriptors, Criteria, and Procedures for Determining Human Response to Noise** 943
Malcolm J. Crocker

81 **Hearing Protection Devices** 967
E. H. Berger and J. G. Casali

82 Determination of Noise Emission of Sound Sources 983
George C. Maling, Jr., and William W. Lang

83 The Generation of Noise in Machinery, Its Control, and the Identification of Noise Sources 991
Malcolm J. Crocker

84 Active Noise Control 1025
P. A. Nelson and S. J. Elliot

85 Noise Sources and Propagation in Ducted Air Distribution Systems 1039
Howard Kingsbury

86 Sound Power Level Predictions for Industrial Machinery 1049
Robert D. Bruce and Charles T. Moritz

87 Airport Noise 1059
Kenneth McK. Eldred

88 Surface Transportation Noise 1073
Robert Hickling

89 Community Response to Environmental Noise 1083
Sanford Fidell and Karl S. Pearsons

Volume Three

PART IX ARCHITECTURAL ACOUSTICS 1093

90 Introduction 1095
A. Harold Marshall and William J. Cavanaugh

91 Sound in Enclosures 1101
K. H. Kuttruff

92 Sound Absorption in Enclosures 1115
David A. Bies and Colin H. Hansen

93 Sound Insulation: Airborne and Impact 1129
A. C. C. Warnock and Wolfgang Fasold

94 Ratings and Descriptors for the Building Acoustical Environment 1161
Gregory C. Tocci

95 Sound Propagation in Work Spaces 1181
Ulrich J. Kurze

96 Acoustical Guidelines for Building Design 1189
Ewart A. Wetherill

97 Noise Control in U.S. Building Codes 1205
Gregory C. Tocci

98 Noise Control for Mechanical and Ventilation Systems 1219
R. M. Hoover and R. H. Keith

PART X ACOUSTICAL SIGNAL PROCESSING 1243

99 Introduction 1245
Andrew F. Seybert

100 Statistical Theory of Acoustic Signals 1249
Allan G. Piersol

101 Practical Considerations in Signal Processing 1261
John C. Burgess

102 Acoustic Holography 1281
Julian D. Maynard

PART XI PHYSIOLOGICAL ACOUSTICS 1291

103 Introduction 1293
Joseph E. Hind

104 Auditory Function 1301
Steven Greenberg

105 Acoustical Characteristics of the Outer Ear 1325
Edgar A. G. Shaw

106 Acoustic Properties of the Middle Ear 1337
William T. Peake and John J. Rosowski

107 Anatomy of the Cochlea and Auditory Nerve 1347
Norma B. Slepecky

108 Cochlear Mechanics and Biophysics 1357
Mario A. Ruggero and Joseph Santos-Sacchi

109 Physiology of the Auditory Nerve 1371
D. O. Kim and K. Parham

110 Anatomy of the Central Auditory Nervous System 1381
Douglas L. Oliver

111 Electrophysiology of the Central Auditory Nervous System 1389
Tom C. T. Yin and Philip H. Smith

112 Biochemistry and Pharmacology of the Auditory System 1401
William F. Sewell

113 Parallel-Hierarchical Processing of Complex Sounds for Specialized Auditory Function 1409
Nobuo Suga

PART XII PSYCHOLOGICAL ACOUSTICS 1419

114 Introduction 1421
David M. Green and Dennis McFadden

115 Auditory Masking 1427
Søren Buus

116 **Frequency Analysis and Pitch Perception** 1447
Brian C. J. Moore

117 **Functions of the Binaural System** 1461
Ervin R. Hafter and Constantine Trahiotis

118 **Loudness** 1481
Bertram Scharf

119 **Effects of High-Intensity Sound** 1497
W. Dixon Ward

120 **Clinical Audiology: An Overview** 1509
Larry E. Humes

121 **Perception of Complex Waveforms** 1521
Charles S. Watson and Gary R. Kidd

122 **Adaptation in the Auditory System** 1535
Paul J. Abbas

123 **Hearing Thresholds** 1545
William A. Yost and Mead C. Killion

Volume Four

PART XIII SPEECH COMMUNICATION 1555

124 **Introduction** 1557
J. L. Flanagan

125 **Models of Speech Production** 1565
Kenneth N. Stevens

126 **Speech Perception** 1579
Joanne L. Miller

127 **Acoustical Analysis of Speech** 1589
Gunnar Fant

128 **Techniques of Speech Coding** 1599
Bishnu S. Atal

129 **Machine Recognition of Speech** 1607
Lawrence R. Rabiner

PART XIV MUSIC AND MUSICAL ACOUSTICS 1615

130 **Introduction** 1617
Thomas D. Rossing

131 **Stringed Instruments: Bowed** 1619
J. Woodhouse

132 **Stringed Instruments: Plucked** 1627
Bernard E. Richardson

133 **Woodwind Instruments** 1635
Neville H. Fletcher

134 **Brass Instruments** 1643
J. M. Bowsher

135 **Percussion Instruments** 1653
Thomas D. Rossing

136 **Pianos and Other Stringed Keyboard Instruments** 1663
Gabriel Weinreich

137 **Pipe and Reed Organs** 1671
Jürgen Meyer

138 **Electronic and Computer Music** 1679
William Morris Hartmann

139 **Human Singing Voice** 1687
Johan Sundberg

PART XV BIOACOUSTICS 1697

140 **Introduction** 1699
Floyd Dunn

141 **Acoustical Characteristics of Biological Media** 1703
Jeffrey C. Bamber

142 **Biological Effects of Ultrasound** 1727
Edwin L. Carstensen

143 **Medical Diagnosis with Acoustics** 1739
K. Kirk Shung and Guy Cloutier

144 **Acoustical Medical Imaging Instrumentation** 1751
James F. Greenleaf

145 **Effects of Vibration and Shock on People** 1761
Henning E. von Gierke

PART XVI ANIMAL BIOACOUSTICS 1781

146 **Introduction** 1783
Arthur N. Popper

147 **Hearing and Sound Perception among Vertebrate Animals** 1789
Richard R. Fay

148 **Insect Bioacoustics** 1799
T. G. Forrest

149 **Amphibians** 1807
Randy Zelick and R. R. Capranica

150 **Bird Acoustics** 1813
Stephen Nowicki

151 **Bats and Echolocation** 1819
James A. Simmons

152 **Nonhuman Primate Vocal Communication** 1823
Marc D. Hauser

153 **Bioacoustics of Marine Vertebrates** 1831
Arthur N. Popper and Peggy L. Edds-Walton

PART XVII ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION 1837

154 **Introduction** 1839
P. V. Brüel, J. Pope, and H. K. Zaveri

155 **Sound Level Meters** 1845
Robert W. Krug

156 **Sound Intensity** 1855
Malcolm J. Crocker and Finn Jacobsen

157 **Calibration of Pressure and Gradient Microphones** 1869
Victor Nedzelnitsky

PART XVIII TRANSDUCERS 1881

158 **Introduction** 1883
Harry B. Miller

159 **Transducer Principles** 1889
Elmer L. Hixson and Ilene J. Busch-Vishniac

160 **Loudspeaker Design** 1903
Bradley M. Starobin

161 **Horns** 1925
Vincent Salmon

162 **Types of Microphones** 1933
Ilene J. Busch-Vishniac and Elmer L. Hixson

163 **Public Address and Sound Reinforcement Systems** 1945
David Lloyd Klepper and Larry S. King

164 **Magnetic Recording Reproducing Systems** 1963
Marvin Camras

165 **Digital Audio** 1967
James W. Beauchamp and Robert C. Maher

166 **Hearing Aid Transducers** 1979
Mead C. Killion

REVIEWERS 1991

INDEX 1993

PART IX

ARCHITECTURAL ACOUSTICS

90

INTRODUCTION

A. HAROLD MARSHALL AND WILLIAM J. CAVANAUGH

1 INTRODUCTION

Architectural acoustics encompasses the acoustical dimensions of the human-built environment. Related expressions that may be used to define this field or parts of it are building acoustics and noise control, architectural engineering (acoustics), architectural science (acoustics), room acoustics, and acoustical design. In this chapter an attempt is made to provide a general framework within which aspects of architectural acoustics can be seen to fit. The term *architectural*, in particular, merits some consideration. The chapter starts with a summary of related handbook sections underlining the cross-disciplinary scope of architectural acoustics. *Foreground* and *background* in the context of architectural design for both noise control and communication lead to a discussion of the relationship between architectural acoustics and human hearing processes and its implications for the preferred properties of rooms for both speech and music. Acoustical art forms as a special case of human communication lead to the expectation that the signal will be enhanced by the room in which it takes place. The idea of "acoustical scale" as a useful way of distinguishing appropriate design concepts is introduced, followed by sections on active architectural acoustics—the interaction of electronics and rooms—and predictive design techniques. Finally there is a brief history of the research in this century upon which the foregoing ideas are based and of the open questions that remain.[1–4]

Because the field includes human subjective response to sounds, a handbook cannot be usefully prescriptive on all points (humans are varied in their preferences), but in spite of the nonlinearity of human responses to different acoustic stimuli at different times, it is possible to see all auditory events as lying on continua subject to the signal processing provided by the binaural auditory tract and the higher functions of the brain. These continua are signal/noise, clarity/unclarity, certainty/uncertainty of localization, and communication/privacy.

2 RELATED PARTS

Other parts of this work are also relevant to problems in architectural acoustics, including the following:

- Part VI (Mechanical Vibration and Shock), particularly Chapter 71 (Vibration Isolation and Damping) and Chapter 75 (Active Vibration Control).
- Part VII (Statistical Methods in Acoustics), notably Chapter 77 (Response Statistics of Rooms).
- Part VIII (Noise Control), especially Chapter 80 (Rating Measures, Descriptors, Criteria, and Procedures for Determining Human Response to Noise).
- Part XII (Psychological Acoustics), particularly Chapters 115 (Auditory Masking), Chapter 116 (Frequency Analysis and Pitch Perception), Chapter 117 (Functions of the Binaural System), and Chapter 118 (Loudness).
- Part XIII (Speech Communication). In particular, Chapter 124 (Introduction) is noteworthy since so many problems in architectural acoustics involve human voice communication as well as speech privacy.
- Part XIV (Music and Musical Acoustics). Chapter 130 (Introduction) is a worthwhile guide to the

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

remaining chapters, 131–139, which cover practically all musical instruments including the human singing voice as well as electronic music and computer music. Nearly every auditorium or church application in architectural acoustics involves music sources that must be considered in order to design optimum environments for their performance and enjoyment.

- Part XVII (Acoustical Measurements and Instrumentation). Since all architectural acoustics applications are based on criteria, guidelines, or standards for the measurement of sound sources and sound transmission paths in buildings, Chapter 154 (Introduction) is particularly relevant.
- Part XVIII (Transducers). Chapter 160 (Loudspeaker Design) and Chapter 163 (Public Address and Sound Reinforcement Systems) deal with many of the applications basic to building spaces.

3 ARCHITECTURAL ACOUSTICS: FOREGROUND OR BACKGROUND

The extent to which the architecture of a building is determined by acoustical considerations depends on the program for the building. Buildings affect the acoustic signal that arrives at the ears, both spectrally and temporally and both within spaces and between them, but in most cases the form of the building is determined by factors other than sound. Acoustics and noise control are then background issues to be dealt with in the choice of functional criteria and the engineering of the fabric of the building and its plant to achieve them. Chapters 91–98 deal with such topics.

It is only when these considerations become the determinants of the built form that the term "architectural" takes its design-centered meaning. Such considerations may take precedence either for noise control, as in the case of a self-protecting, naturally ventilated building designed to protect its opening windows by turning away from freeway noise, or, more often, for the acoustical design of spaces in which communication is important. For such buildings and rooms the acoustical considerations are foreground, in that if they are not correctly included in the design thinking from the start, there may not be a solution available that is not in conflict with the architectural intentions for the space. The chances of achieving excellence in the quality of either the building or the communication within it are not great in this case. "Correct" inclusion implies that the design is based on the realities of the physics of sound and the properties of human hearing and includes quantitative prediction of the sought-for outcome at the design stage of the project.

4 ARCHITECTURAL ACOUSTICS AND COMMUNICATION

The idea of communication underlies architectural acoustics and gives rise to several of its units of measurement. The signal-to-noise ratio, for example, is the fundamental measure for all auditory tasks. A derivative measure, the articulation index (AI), takes the specifics of speech intelligibility, masking, and vocal power to produce a measure as useful for rating privacy as it is for ease of communication with speech. Many of the other measures for the noise environment are single sided in that they rate only difficulty of communication against noise levels but in each case speech communication is the yardstick. Examples of such measures in common use are the noise criteria family of curves (NC, PNC, RC, and NCB, see Chapter 80) and the noise-rating numbers (NR). Inspection of the shape of these curves and the rating basis for partition performance (sound transmission class) reveals the common factor between them in the intelligibility of transmitted speech.

In a fundamental way that is not recognized in these units, concepts that are useful take account of the fact of binaural hearing. It is only in extreme situations for speech communication such as a poor signal-to-noise ratio or of several competing sources that the benefits of a binaural system become apparent. Then humans are able to achieve a high degree of subjective acoustic imaging that enables localization of the source, signal enhancement, and its converse, discrimination against the competing sources. Further, sound from the source (a person or a loudspeaker), reflected to arrive within about 50 ms of the direct sound, is usefully added by human hearing to it, effectively increasing its level and clarity—though the extent to which this is a binaural process clearly depends on the arrival direction of the reflected energy. Several useful measures for the ratio of early to late energy are good predictors of the adequacy of speech communication, and these are all single-channel measures (D50, C80, STI). Placing reflecting surfaces to take advantage of reflections is as fundamental to the design of excellent rooms for speech communication as the more conventional control of echo and reverberation. Strong, late reflections and excessive reverberation are as destructive of clarity in speech communication as the early reflections are useful. The integration function for reflections of speech and the threshold for the occurrence of echo as a function of the level and delay of reflections have been known since the 1960s.

The binaural properties of hearing and its integration function are even more significant in the design of rooms for the performance of the other principal human communication, music. There, however, the variability of the signal and the variety of listener preference require several measures on the sound fields, and there is not yet complete unanimity as to what these should be. The problem arises in that musical style ranges from the cerebral to the emotional and is appreciated differently according to the aptitudes and sensibilities of the audience. Clarity, and with it the precise imaging of the source, which is a universally desired property of a speech communication channel, is undesired on the emotional axis of music. Thus some weakening of localization is preferred, especially with music of the Romantic genre, which currently dominates the Western concert programs. That weakening is achieved by decreasing the early–late ratio and increasing the relative strength of lateral reflections. The former increases reverberance (nondirectional), while the latter increases apparent source width and the sense of envelopment in the sound (weaker localization).

The common ground among the competing music room measures is that strength and reverberance are essential, and these are measured monaurally. Various measures for the early–late ratio are current, also measured with a single channel, among them the early–late energy ratio C and the center time T_c. The lateral reflected sound is measured either by some version of the interaural correlation coefficient (IACC) or directly by the lateral-to-frontal energy ratio in the arriving sound. It now appears that source broadening depends on the early lateral sound while the sense of envelopment is highly correlated with the late (after 80 ms) lateral strength. Other contenders as contributors of excellence are a spectral measure and the delay of the first strong reflections related to the autocorrelation function of the music. The latter means that the rate of change of the music determines the preference for the delay of the reflections after the direct sound.

5 ACOUSTICAL ART FORMS

It is plain, then, that enclosures such as rooms affect the signal transmitted acoustically through them in ways that may be regarded as either positive or negative. Architectural acoustics seeks to address these issues so that the outcomes are positive in terms of the functions for which the room is intended. The most demanding level of acoustical communication occurs in the art forms—theater, poetry, music—and rooms for these make correspondingly high demands for acoustical excellence. The acoustical attributes of enclosure have come in this century to be recognized as potential enhancements of the signal, particularly for music and less obviously also for speech communication. That this is a recent phenomenon is illustrated by the preference studies of the past two decades and the fact that in 1941 F. R. Watson could write that the objective of acoustical design was to make room conditions as similar to outdoors as possible.

6 ACOUSTICAL SCALE

The audible effects of rooms are markedly different in small spaces from those observed in large spaces. The concept of "acoustical scale" is helpful in accounting for these differences. Acoustical scale relates size (of rooms or building elements) to wavelength and so to frequency. It follows that a given room may behave as a small room up to some frequency and above it as a large room. Small-room response to sound is dominated by modal behavior. In large rooms, modes are so numerous that reasonably reliable statements about average energy density, decay rates, and reverberation times and the like can be made. The useful notion of sound "rays" is also available to track early reflected sound from surfaces that are large relative to the wavelength. The transition frequency between the two types of behavior, called the Schroeder frequency, fs, after its proposer M. R. Schroeder, is given by $fs = 2000/TV$, where T is the reverberation time in seconds and V is the room volume in cubic metres. Below this frequency it makes little sense to calculate the reverberation time.

7 ACTIVE ARCHITECTURAL ACOUSTICS

For many years electronic devices have been available to assist the natural acoustical properties of enclosure. These consisted initially of reinforcement and amplification systems with, from the mid-1950s, a variety of artificial reverberation devices. Synthetic "reflections" could be achieved with tape-delay devices but with all the disadvantages inherent in a mechanical tape transport system. The advent of digital signal processing has overcome these problems, and it seems certain that more versatile and sophisticated systems will continue to develop. In parallel with these developments active control of noise using cancellation techniques has become a reality in fields as diverse as automotive design and dentistry, with a considerable application in architectural acoustics. Again one sees a continuum, in this case device based, that extends from the control of noise to the enhancement of communication.

8 PREDICTIVE TECHNIQUES DURING DESIGN

While some quantities may be calculated easily during design,[5] others such as energy fractions in rooms dominated by the early reflection sequence are less accessible. The advent of low-cost digital computers has given rise to numerous programs for these predictions from digitized representations of the space, allocating absorption and diffusion coefficients to the various surfaces and either tracing rays from a source or projecting image sources in the surfaces. Depending on the complexity of the room, measures derived from the impulse response may be obtained by this means during the design stage. The more complex the room, especially if it includes curved surfaces that are difficult to digitize, the greater the uncertainty in the results. A third, hybrid method makes use of the advantages of digital data acquisition in a scale physical model of the room, compensating for the absorption of the real air conditions numerically during the signal processing. The same process is then available at full size during commissioning for measurements on the finished room. Scales commonly chosen for the model range from 1:50 up to 1:10, the only advantage of the latter being the availability of scaled music reproduction on a multispeed tape recorder.

Recent developments in auralization may obviate even that advantage.

9 HISTORY OF RESEARCH IN ARCHITECTURAL ACOUSTICS

Development of knowledge in architectural acoustics falls into three phases. Starting with the pioneering work of Wallace Clement Sabine and leading in the late years of the nineteenth century to the well-known relationship $T = kVA$, there was a period of consolidation during which all acoustical design was based solely on the control of reverberation time. Sabine's discovery was so far-reaching and such an advance on the previous state of knowledge that it appeared that all architectural acoustical issues were resolved in it. The subtleties of the perception of reflections as components of reverberation processes did not appear as important factors for nearly 50 years. Indeed, the preoccupation of the Western world with two world wars and the depression of the 1930's provided limited scope for major communications spaces during this period. The second phase, starting after World War II, addressed the significance of the early reflections, to account among other matters for the observed fact that measured reverberation time did not always represent the subjective impression adequately. Further, the notion that the gratifying listening experience in concert halls depends on the temporal details of the early reflections was significantly extended by the discovery that lateral reflections made an important contribution to the sense of spaciousness found in the most preferred halls. That is, both the spatial and temporal properties of the early reflections must be controlled in the design in addition to the reverberation time. The shape of the room, ignored in the Sabine relationship, must be taken into account in acoustical design. The integration time and absolute threshold of perceptibility for the early reflections and the threshold for echo for speech and other signals became the focus of research. Even more significant was the application of factor analysis to rigorously designed preference experiments, the demonstration that preference is multidimensional, and the isolation of four principal subjective factors determining it. Objective correlates for these factors were proposed: reverberation time (RT) and early decay time (EDT) for reverberance, the total level referred to the level of the direct sound at 10 m from the source (G) for "strength", the 80-ms energy fraction (C80) and related measures such as Cremer's center time T_c for clarity; and either the 80-ms lateral fraction (LF) or the IACC for spatial impression. Space here precludes naming the key authors during this period, and the reader is referred to Refs. 6–9 for a comprehensive bibliography.

In this century the third phase of architectural acoustics research has centered on refinements of listener preference criteria.

Open areas of research are changing preference frame and changing art forms. The third phase of research is on-going and will remain so especially for the art forms, as the media themselves change the preference frame within which judgments of excellence are made. One has only to see the way views have changed in 50 years since Floyd Watson's dictum in 1941 to guess that listening habits follow developments in equipment. Perhaps of even greater significance will be the selection of the classical concert repertoire, which has been dominated by the Romantic composers of recent centuries—only a blip in the history of musical communication. Finally the art forms themselves change, again following availability of technical development. The last two decades of the twentieth century have seen the widespread availability of digital computers that lead to entirely new possibilities in what constitutes music, including interaction of the visual and audible. This trend is accelerating without any slackening in the popularity of other forms of auditory entertainment.

In this introductory chapter a rationale for the concerns of architectural acoustics has been sought in the way human binaural hearing deals with sounds. The fact that, on the basis of the sheer weight of practice time, noise control issues[10] in this field may appear more important than communication issues has tended to obscure this rationale. Noise that intrudes, annoys, interferes, and to which humans find difficulty in adaptation is, according to the picture presented here, precisely what is inherent in music. The acoustical properties of enclosure that accentuate the noise enhance the art.

REFERENCES

1. F. V. Hunt, *Origins in Acoustics* (with foreword by R. E. Apfel); American Institute of Physics, Woodbury, NY, 1992, originally published 1978, Yale University Press.
2. J. W. Kopec, *The Sabines of Riverbank.* Peninsular, Woodbury, NY, 1997.
3. T. D. Northwood, Ed., *Architectural Acoustics*, Vol. 10, Benchmark Papers in Acoustics Series, Dowden, Hutchinson and Ross, Stroudsburg, PA, 1977.
4. W. C. Sabine, *Collected Papers on Acoustics*, Peninsular, Woodbury, NY, 1994, originally published 1921 Harvard University Press and 1964 Dover.
5. V. O. Knudsen and C. M. Harris, *Acoustical Designing in Architecture*, American Institute of Physics, Woodbury, NY, 1980, originally published 1950, Wiley, New York.
6. M. Barron, *Auditorium Acoustics and Architectural Design*, E & FN Spon Publisher, an imprint of Chapman & Hall, London and New York, 1993.
7. L. L. Beranek, *Concert and Opera Halls ... How They Sound*, American Institute of Physics, Woodbury, NY, 1996; originally published, *Music, Acoustics and Architecture*, 1962, Wiley, New York.
8. L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics*, Vols. I and II, Applied Science Publishers, London and New York, 1978 (translated by T. J. Schultz).
9. M. D. Egan, *Architectural Acoustics*, McGraw-Hill, New York, 1988.
10. C. M. Harris, Ed. *Noise Control in Buildings*, McGraw-Hill, New York, 1994.

91

SOUND IN ENCLOSURES

H. KUTTRUFF

1 INTRODUCTION

This chapter deals with the physics of sound propagation in enclosures, that is, in spaces filled with a homogeneous fluid—for instance, with air—that are completely surrounded by a passive boundary.

We restrict the discussion to enclosures with dimensions larger than all sound wavelengths of interest. Although most results may be applied to small spaces such as living rooms or the interior of passenger cars, our main interest is the acoustical behavior of large rooms such as performance halls or noise propagation in all kind of working environments. In such cases, the sound field is made up of many components whether we try to describe it in terms of normal modes or of sound rays or particles.

This chapter presents three different approaches to describe sound fields. The first one involves the acoustic wave equation and solutions of it. This procedure yields the physically correct theory of sound in enclosures. Although it is not particularly well suited for solving practical problems, it is the basis of a more than superficial understanding of sound fields in enclosures. A more illustrative description is obtained by restricting the discussion to the propagation and distribution of sound energy. This is effected by introducing sound rays and image sources that yield a simple method of computing the distribution of sound energy.

For many practical purposes it is reasonable to compute the sound field in an enclosure only to that degree of accuracy needed to evaluate certain characteristic parameters of it, such as the stationary energy density and the reverberation time. Closed expressions are obtained by considering energetic averages rather than the energies of individual sound rays or sound particles.

2 WAVE ACOUSTICS

2.1 Normal Modes and Eigenfrequencies

The basis of a physically exact description of sound propagation in an enclosure is the acoustic wave equation. If dissipation in the medium is neglected and if harmonic wave motions with angular frequency ω are considered, the wave equation transforms into the Helmholtz equation

$$\nabla^2 p + k^2 p = 0. \tag{1}$$

In this differential equation p is the complex sound pressure amplitude and $k = \omega/c$ denotes the (angular) wavenumber; c is the sound velocity.

The effect of the boundary on the sound field can be accounted for by imposing certain boundary conditions on the possible solutions of Eq. (1). If the boundary is locally reacting to the sound field, these conditions can be expressed in terms of its complex wall impedance Z, which is usually a function of the sound frequency ω and the location:

$$\omega\rho_0 p + iZ\frac{\partial p}{\partial n} = 0 \quad \text{along the boundary.} \tag{2}$$

Here, $\partial p/\partial n$ is the component of the sound pressure gra-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

dient in the direction of the outward-pointing wall normal and ρ_0 is the density of the medium.

The Helmholtz equation (1) supplemented by the boundary conditions [e.g., as expressed by Eq. (2)] can only be solved for discrete values k_n of the wavenumber, the so-called eigenvalues. (The subscript stands for three integers, e.g., n_x, n_y, n_z.). They are related to *allowed frequencies*, or *eigenfrequencies*, $f_n = ck_n/2\pi$, or when expressed as angular frequencies, $\omega_n = ck_n$. Each eigenvalue is associated with an *eigenfunction*, or *characteristic function*, $p_n(\mathbf{r})$, the vector $\mathbf{r}$ symbolizing the space coordinates. (Strictly speaking, more than one eigenfunction may belong to one eigenvalue. However, this case, known as *degeneracy*, is of no importance in practical situations.) Another common expression for an eigenfunction is *normal mode*. It can be thought of as a three-dimensional standing wave the nodes of which are situated on nodal surfaces. Of course, the distinctness of these standing waves depends on the reflective properties of the boundary, in an anechoic room, for instance, there will be mainly free-wave propagation superimposed by faint standing waves.

If the wall impedance is complex, which is usually the case, the eigenvalues are also complex: $k_n = (\omega_n + i\delta_n)/c$. Accordingly, the time dependence of the sound pressure is $\exp(i\omega_n t - \delta_n t)$, indicating that the imaginary parts can be interpreted as damping constants. In this case the standing waves are not persistent but will die out unless a sound source will continuously support the energy absorbed by the boundary. This is the basis of sound decay or reverberation, which will be dealt with later in more detail.

If the boundary of the enclosure is not locally reacting, that is, if waves propagating in or behind the walls can interact with the sound field inside the considered enclosure, the boundary conditions cannot be formulated by such a general expression as Eq. (2). In this case, which has been considered by Pan and Bies,[1] the normal modes of the enclosure may be coupled to modes within or outside the boundary.

Once the normal modes of an enclosure are known, its acoustical response to any excitation by a sound source can be formally expressed by them. However, closed expressions for the eigenfunctions and eigenfrequencies can be worked out only for enclosures with simple geometry and boundary conditions (see below). In most practical cases where the shape of the room and the acoustical properties of its walls are complex, it may turn out quite difficult to determine even a moderate number of normal modes. In general, one has to resort to numerical methods such as the finite-element method, and even then only small enclosures and/or the low-frequency range where the density of eigenfrequencies is relatively small can be treated in this way. For large rooms a geometric analysis based on sound rays may be more useful (see Section 3).

2.2 Example: Normal Modes of a Rectangular Enclosure

Let us consider a rectangular room bounded by three pairs of parallel planes that are perpendicular to each other and to the axis of a Cartesian coordinate system (see Fig. 1) the origin of which coincides with one of the room corners. The side lengths of this box are L_x, L_y, and L_z.

Then Eq. (1) reads

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} + k^2 p = 0. \tag{3}$$

Furthermore, we assume that the walls are rigid, that is,

$$\frac{\partial p}{\partial x} = 0 \quad \text{for } x = 0,\ x = L_x. \tag{4}$$

Two similar boundary conditions hold for the y- and z-directions.

The solutions of Eq. (3) satisfying these conditions are given by

$$p_{n_x n_y n_z}(x, y, z) = A \cos\left(\frac{n_x \pi x}{L_x}\right) \cos\left(\frac{n_y \pi y}{L_y}\right) \cdot \cos\left(\frac{n_z \pi z}{L_z}\right) \tag{5}$$

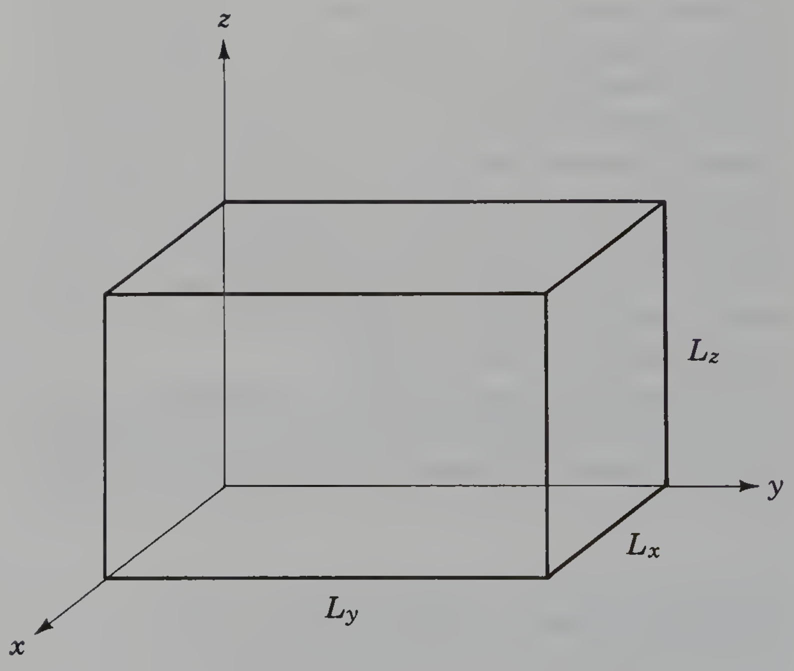


Fig. 1 Dimensions of a rectangular enclosure.

with an arbitrary constant A; the eigenfrequencies are

$$f_{n_x n_y n_z} = \frac{c}{2}\left[\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2\right]^{1/2} \tag{6}$$

with n_x, n_y, and n_z denoting integers including zero. The normal mode is an *axial* one if two of these indices vanish; it is formed by successive sound reflections between one pair of parallel walls. If two pairs of walls are involved, we speak of a *tangential* mode, for which one of the three indices is zero. All remaining modes are *oblique modes*.

Figure 2 shows the curves of constant sound pressure at a given moment in the plane $z = 0$ for $n_x = 3$, $n_y = 3$, and arbitrary n_z. Within adjacent rectangles the pressure has opposite signs at the moment considered; half a period later all signs will have changed. Along the dotted lines the sound pressure is zero for any time. These lines are the intersection of the plane $z = 0$ with two systems of equidistant nodal planes; a third system is parallel to $z = 0$. The numbers of nodal planes perpendicular to the x-, y-, and z-axis are given by the indices n_x, n_y, and n_z, respectively.

The number of eigenfrequencies within the frequency range from 0 to an upper limit f can be shown to be given approximately by

$$N_f \approx \frac{4\pi}{3} V\left(\frac{f}{c}\right)^3 + \frac{\pi}{4} S\left(\frac{f}{c}\right)^2 + \frac{L}{8}\frac{f}{c}. \tag{7}$$

In this expression V is the volume of the room and S is the area of its boundary; $L = 4(L_x + L_y + L_z)$ is the sum of the lengths of all edges of the room. It is noteworthy that in the limit of very high frequencies the first term of this formula is valid for any enclosure, that is, not only for rectangular enclosures with rigid walls. The same holds for the average spacing of adjacent eigenfrequencies, which is obtained by differentiating the first term of Eq. (7) with respect to the angular frequency:

$$\langle \delta f \rangle \approx \frac{c^3}{4\pi V f^2}. \tag{8}$$

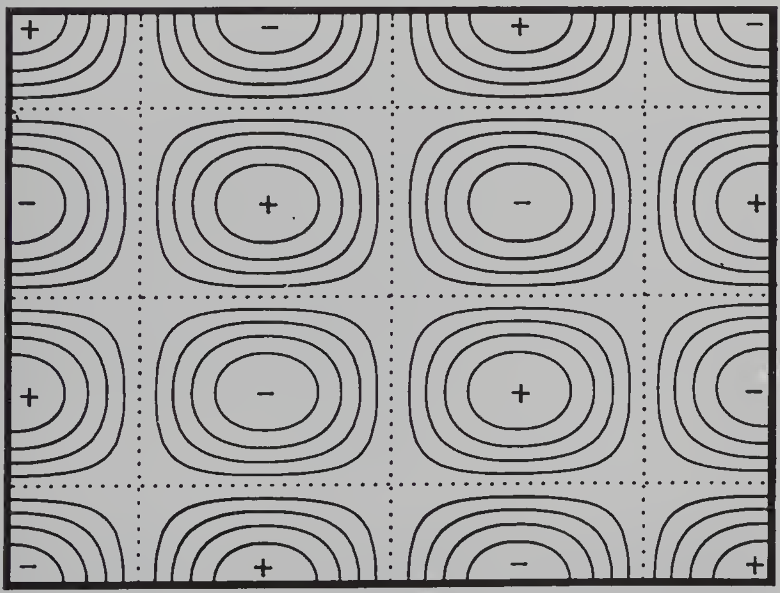


Fig. 2 Normal mode in a rectangular enclosure (see Fig. 1): curves of equal sound pressure amplitude in a plane perpendicular to the z-axis ($n_x = 3, n_y = 3$). (From H. Kuttruff, *Room Acoustics*, Chapman & Hall, London, 1991.)

Equation (7) shows that in larger rooms the number of eigenfrequencies needed to cover the audio frequency range may be quite large. This is another reason why the eigenfunction concept is of little use for solving problems in practical room acoustics.

2.3 Steady-State Sound Field

If all eigenvalues $k_n = (\omega_n + i\delta_n)/c$ and the associated eigenfunctions $p_n(\mathbf{r})$ are known for a given enclosure, the complex sound pressure amplitude in any inner point $\mathbf{r}$ can be expressed in terms of them. Let us suppose an enclosure the normal modes of which have been determined in some way and that is excited by a point source at a location $\mathbf{r}'$ operating at an angular frequency ω. Then, under the assumption $\omega_n \ll i\delta_n$, the complex sound pressure amplitude in a point $\mathbf{r}$ is obtained as

$$p_n(\mathbf{r}) = C\sum_{n=0}^{\infty} \frac{p_n(\mathbf{r})p(\mathbf{r}')}{K_n(\omega^2 - \omega_n^2 - 2i\delta_n\omega_n)}, \tag{9}$$

where the constant C contains the strength of the sound source; K_n is a normalization constant: For rectangular rooms it is 1/2 for axial modes, 1/4 for tangential modes, and 1/8 for oblique modes. Each term of the sum in Eq. (9) reaches its maximum when the driving frequency ω is close to the corresponding angular eigenfrequency ω_n, that is, if it represents a resonance of the enclosure.

Concerning the frequency dependence of the pressure amplitude, two limiting cases have to be distinguished: Suppose the average spacing of eigenfrequencies after Eq. (8) is substantially smaller than the average resonance half-width $\langle \Delta f \rangle = \langle \delta \rangle / 2\pi$, with $\langle \delta \rangle$ denoting the average of the damping constants δ_n. Then the transfer function of the enclosure consists of an irregular succession of well-separated resonance curves. In the opposite case, which is typical for large rooms, several or even many eigenfrequencies are situated within the average half-width. Then the corresponding resonance curves show considerable overlap. Hence several or many terms

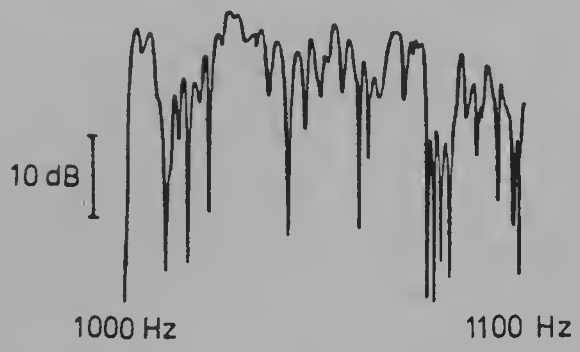


Fig. 3 Measured frequency curve of an enclosure in the range of 1000–1100 Hz. (From H. Kuttruff, *Room Acoustics*, Chapman & Hall, London, 1991.)

of the sum in Eq. (9) contribute to the total sound pressure even if the room is excited with a sine signal. Since these contributions are superimposed with virtually random phases, the sound pressure can be considered as randomly distributed. The condition for this case is[2]

$$f > \frac{5000}{(V\langle\delta\rangle)^{1/2}} \approx 2000\sqrt{\frac{T}{L}}, \qquad (10)$$

where T is the so-called reverberation time of the enclosure, which will be defined in the next section. As an example, Fig. 3 represents the steady-state sound pressure level measured in a particular room as a function of the driving frequency; the recording covers only a relatively narrow frequency interval. The irregular shape of this curve is typical for all such *frequency curves* of an enclosure, no matter where the sound source or the receiving point is located, and their general appearance is the same even for virtually all enclosures.[3] This means that no information on acoustical properties or peculiarities of a performance hall can be derived from the statistical quantities of such recordings.

The following facts hold under the assumption that the distance between the sound source and the observation point is much larger than the so-called reverberation distance [see Eq. (33)]. Then the magnitudes $|p|$ of the sound pressures encountered in any large enclosure follow a Rayleigh distribution: Let q be the magnitude $|p|$ divided by its average; then the probability that this quantity lies within the interval from q to $q+dq$ is given by

$$P(q)\,dq = \frac{\pi}{2}\exp\left(-\frac{\pi}{2}q^2\right) q\,dq \qquad (11)$$

(see Fig. 4). From this distribution density it can be concluded that the sound pressure level fluctuates within a range of about 10 dB on the average. Furthermore, the mean spacing between the maxima of a frequency curve is approximately[2]

$$\langle\delta f_{\max}\rangle \approx \frac{\langle\delta_n\rangle}{\sqrt{3}} \approx \frac{4}{T}. \qquad (12)$$

As before, T is the reverberation time. When the driving frequency is continuously altered, not only the amplitude of the sound pressure in the receiving point **r** but also its phase varies in an irregular way. The average phase variation per hertz is[4]

$$\left\langle \frac{d\Phi}{df} \right\rangle \approx \frac{\pi}{\langle\delta\rangle} \approx \frac{T}{2.2}. \qquad (13)$$

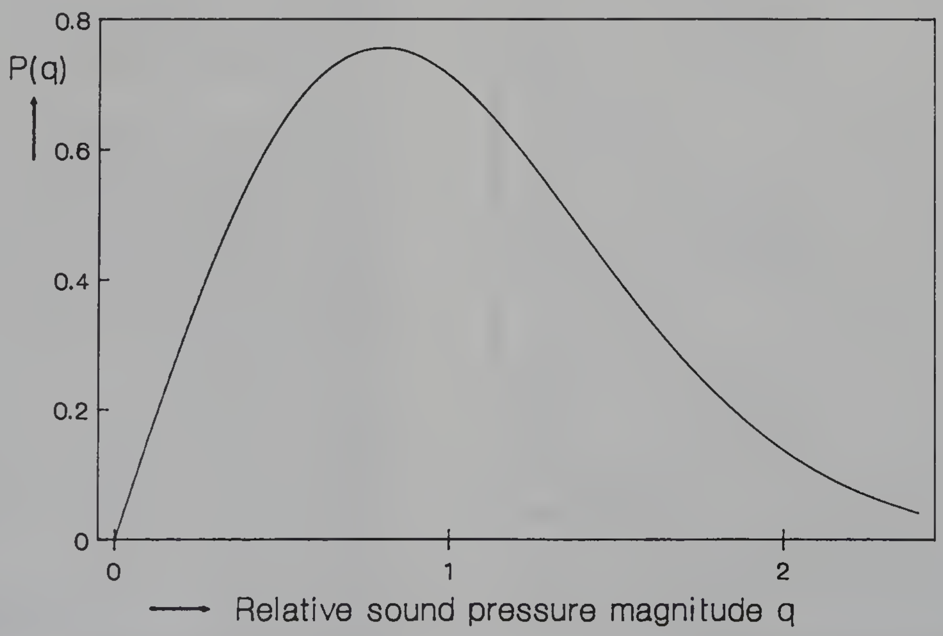


Fig. 4 Distribution of sound pressure amplitudes in a large enclosure (Rayleigh distribution).

If not the driving frequency but the location **r** of the receiving point is continuously altered while the frequency is kept constant, similar fluctuations of the sound pressure amplitudes occur. Since the samples of the sound pressure are taken from the same ensemble as before when the frequency was varied, the distribution of the sound pressure magnitudes is again given by Eq. (11).

2.4 Transient Response: Reverberation

As noted in Section 2.1, any normal mode of an enclosure with complex wall impedance will decay unless a sound source compensates for the wall losses. Therefore, if the room is excited by some sound source that is abruptly stopped at time $t = 0$, all excited normal modes will die out with their individual damping constants δ_n. Accordingly, if we assume $\delta_n \ll \omega_n$, the sound pressure in any point can be expressed by

$$h(t) = \sum_{n=0}^{\infty} a_n e^{-\delta_n t} e^{i\omega_n t} \quad \text{for } t \geq 0. \tag{14}$$

The complex coefficients a_n contain the eigenfunctions p_n and hence the spatial dependence of the sound pressure. Furthermore, they depend on the location of the sound source and the signal it emitted in $t < 0$. Equation (14) describes what is called *reverberation* in room acoustics.

Very often the damping constants δ_n do not differ too much from each other and may be therefore replaced without much error by their mean value $\langle\delta\rangle$. Then the energy density in a certain point of the sound field will decay at a uniform rate:

$$u(t) = u_0 e^{-2\langle\delta\rangle t} \quad \text{for } t \geq 0. \tag{15}$$

In room acoustics, the decay rate is usually characterized not by the mean damping constant but by the *reverberation time*, or the *decay time*, T, defined as the time interval (see Fig. 5) in which the sound energy or energy density drops to one millionth of its initial value. This is tantamount to the requirement that the sound pressure level falls down by 60 dB. From Eq. (15) it follows that

$$T = \frac{3 \ln 10}{\langle\delta\rangle} = \frac{6.91}{\langle\delta\rangle}. \tag{16}$$

The reverberation time is one of the most important quantities in room acoustics. It can be measured with good accuracy, and the available formulas predict it with reasonable accuracy. More will be said about this matter in Section 4.3.

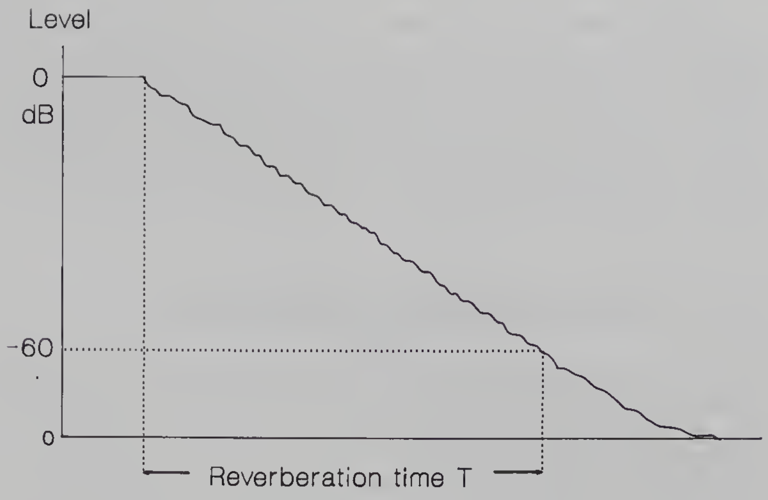


Fig. 5 Sound decay and definition of the reverberation time.

3 GEOMETRIC ACOUSTICS

Although the formal solution of the wave equation as outlined in the preceding section yields many interesting and important general results, its practical application turns out to be too complicated, except for small enclosures and low frequencies.

We arrive at a much simpler, although less rigorous, description if we restrict the discussion to mere energy propagation. This means that all effects involving phase differences such as interference or refraction are neglected. This is admissible if the sound signals of interest are not sinusoids or other signals with small-frequency bandwidth but are composed of many spectral components covering a wide frequency range. Then it can be assumed that constructive or destructive phase effects cancel each other when two or more sound field components are superimposed at a point, and the total energy in the considered point is simply obtained by adding their energies. Components with this property are often referred to as *mutually incoherent*.

Now for the discussion of propagation of sound energy, the wave concept would lead to unnecessary complications. Instead, it is more convenient and illustrative to employ the concept of sound rays or bundles of sound rays. A sound ray can be thought of as a small sector that has been cut out of a spherical wave and subtends a vanishingly small solid angle. These sound rays, which are straight lines in a homogeneous medium, are the paths along which the sound energy travels with constant velocity, and from their definition it follows that this energy does not depend on the distance it has traveled provided the medium is free of dissipation. The intensity within a sound ray, however, decreases proportionally to l/r^2, where r denotes the distance from the sound source. In this picture, uniform radiation of sound from a point source is represented as a diverging bundle of rays originating from it.

Furthermore, we assume that all sound-reflecting objects, in particular all walls of a room, are large compared to the acoustic wavelength. Then the reflection of sound obeys the same laws as those known from optics; that is, if a reflecting surface is smooth, the usual reflection law can be applied.

3.1 Sound Reflection: Image Sources

When a sound ray falls upon a plane and smooth wall of infinite extension, it will be at least partially reflected from it. The law of specular reflection states that the angle under which the reflected ray leaves the wall equals the angle of incidence where both angles are measured against the wall normal (see Fig. 6*a*) and that the incident ray, the reflected ray, and the wall normal are lying in a plane. This law can also be applied to walls of finite size if its dimensions are much larger than wavelengths of all spectral components involved. Consequently, any diffraction effects from, for example, wall edges are neglected in this "high-frequency" limit of sound propagation.

If the incident ray is emitted by a point source, the reflected ray can be thought of as originating from a virtual sound source that is the mirror image of the original source with respect to the reflecting wall (see Fig. 6*b*). This secondary source, which is assumed to emit the same signal as the original one, will be referred to as the *image source*. It is particularly useful for constructing the reflection of a ray bundle from a plane wall portion or for connecting a given sound source and receiving point by a path that includes wall reflections. Its full potential is developed, however, in the discussion of sound propagation in enclosures.

The fact that only a fraction of the incident sound energy will be reflected from the wall can be accounted for by its absorption coefficient α, that is, the fraction of the incident sound energy absorbed by the wall. Accordingly, the reflection process reduces the energy of the sound ray by the factor $1 - \alpha$. This is tantamount to operating the image source at a power reduced by this factor. In this way also frequency-dependent absorption coefficients can be taken into regard; the dependence of the absorption coefficient on the direction of sound incidence could be accounted for by modifying the directivity of the image source. If the sound source itself has a certain directivity, the same but symmetrically inverted directivity pattern must be assigned to the image source.

The law of specular reflection is valid for walls that are plane and smooth. It can also be applied to curved walls, provided their radius of curvature is very large compared to the wavelength. However, the concept of image sources fails in this case. Instead, the reflection of each ray has to be treated separately by constructing the reflected ray path with respect to the tangential plane in the point where the incident ray hits the wall.

If there are two or more plane reflecting walls, an image source can be attributed to each of them. Multiple reflections can be accounted for by higher order image sources that are constructed by mirroring the previously obtained images. This is illustrated by Fig. 7, which depicts an edge formed by two adjacent plane walls. In addition to the two first-order image sources A_1' and A_2', there are two images of second order, A_1'' and A_2''. Viewed from the receiving point, an image source may turn out to be "invisible," as is the case for A_2''. This happens whenever a ray intersects the plane of a wall outside its physical boundary.

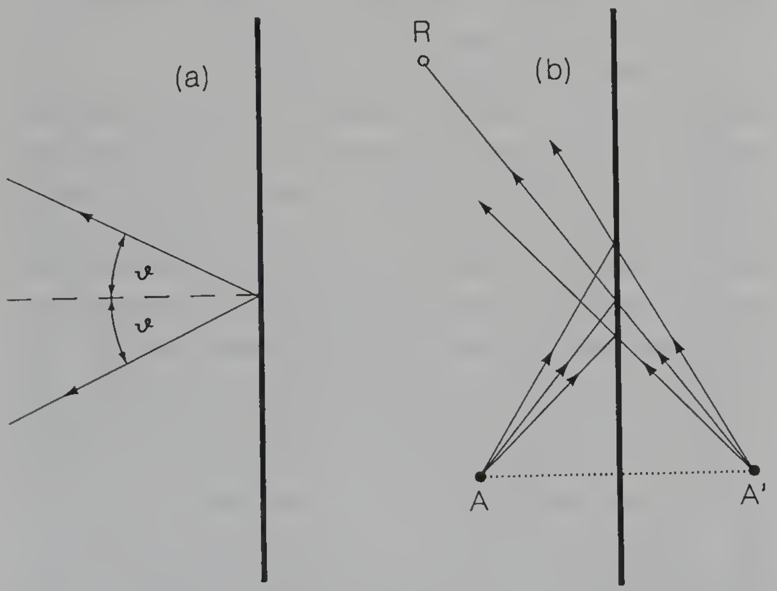


Fig. 6 Specular sound reflection from a plane wall for (*a*) one ray and (*b*) several rays: *A*, original sound source; *A′*, image source.

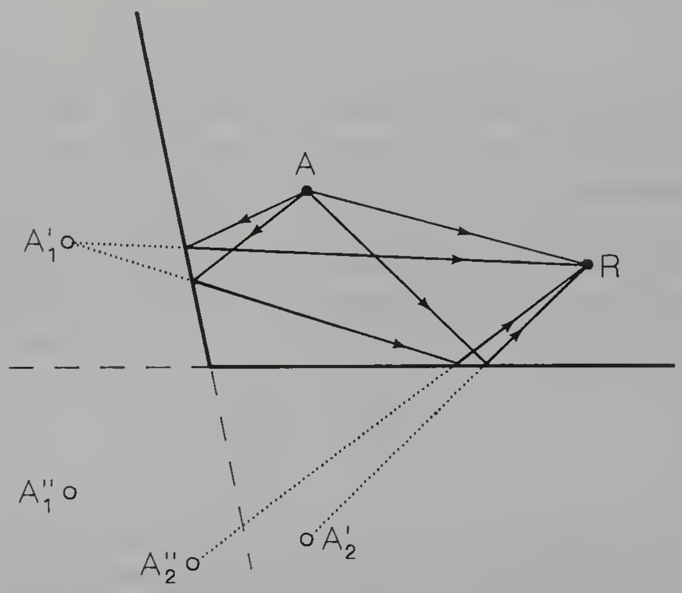


Fig. 7 Sound reflections from an edge formed by two adjacent walls: *A*, original sound source; *A′*, first-order image sources; *A″*, second-order image sources.

For a space that is completely enclosed by plane walls, this process is repeated over and over, leading to images of increasing order. If the enclosure is made up of N plane walls, the total number of image sources of up to order m is

$$n_m = N\,\frac{(N-1)^m - 1}{N-2}. \tag{17}$$

These images form a three-dimensional pattern that depends on the geometry of the enclosure. However, most of these image sources are invisible from the given observation point. Therefore, it has to be checked which of them will really contribute to the total energy in a given point and which do not.[5,6] As before, the absorption coefficients of the walls are taken into account by attributing properly reduced powers to the image sources. Then, for discussing the energy distribution within an enclosure with plane and smooth walls, the boundary itself is no longer needed, and the energy density (or the sound pressure) in a certain point is obtained just by adding the contributions of all visible image sources provided they are not too faint. The effect of air attenuation is easily included if desired.

In this way, not only the steady-state energy density but also transient processes can be calculated. Suppose the original sound source would produce the energy density $u(t)$ in some point of an unbounded space. Under the assumption of frequency-independent absorption, each image source produces a weaker replica of this energy signal that will be received with some delay τ_n, depending on its distance. Thus the received energy signal is

$$u'(t) = \sum_n a_n u(t - \tau_n). \tag{18}$$

If the signal emitted by the sound source is a very short impulse represented by a Dirac function $\delta(t)$, the result of this superposition is the energetic *impulse response* of the enclosure with respect to a particular source–receiver arrangement. A schematic example is shown in Fig. 8. Each reflection is represented by a vertical line the length of which indicates its relative energy, whereas the abscissa is its delay time τ_n with respect to the direct sound, which is represented by the first line. Measured impulse responses differ from this simple scheme since the acoustical properties of real walls usually depend on the sound frequency.

It should be mentioned that the image source model is not limited to the treatment of energy propagation but can be used as well to obtain the sound pressure in a receiving point for any source signal, at least in the limit of high frequencies.

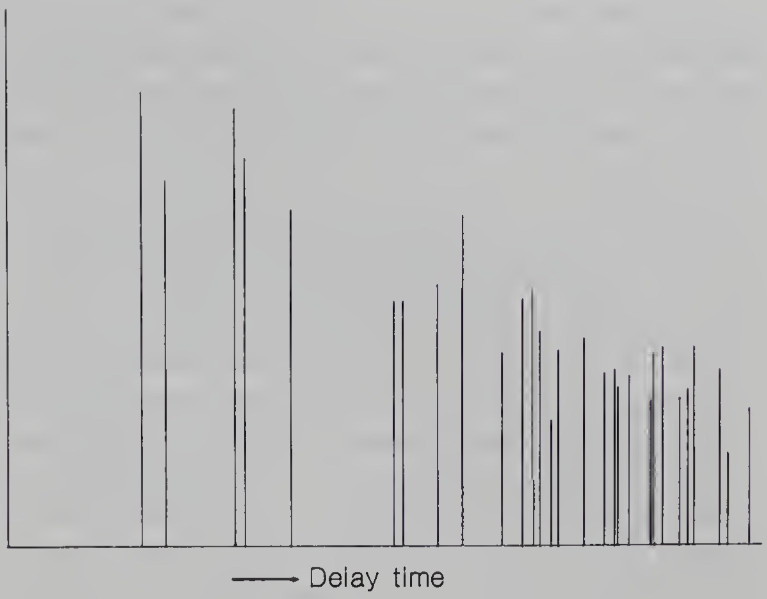


Fig. 8 Energetic impulse response of an enclosure (schematically).

3.2 Two Examples: Flat Room and Rectangular Room

For certain enclosures with simple geometry, all image sources are visible and are arranged in a regular pattern. Accordingly, the total sound energy density in a given point can be represented by a closed formula.

As a first example, we consider the space bounded by two parallel infinite planes with distance a, as depicted in Fig. 9. This enclosure can be considered as the model of many factory spaces or open-plan bureaus the height of which is small compared to their lateral dimensions. Since most points are far from the side walls, the contributions of the latter can be neglected. The image sources

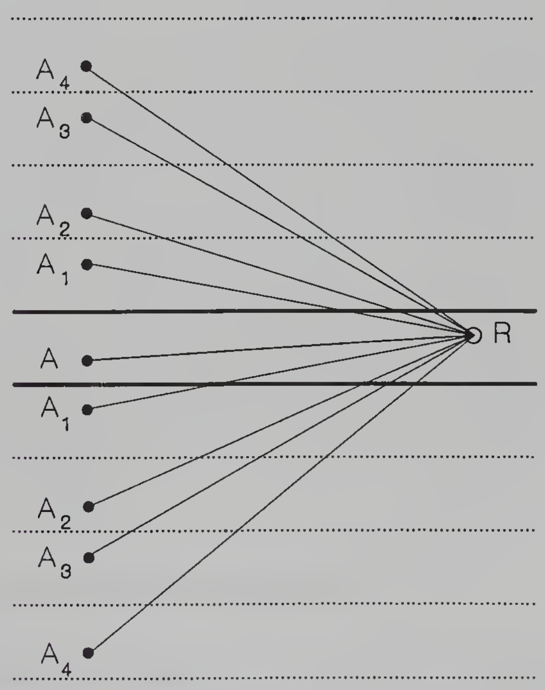


Fig. 9 Flat room and its image sources.

are regularly arranged along a vertical line; if the primary source lies exactly in the middle between the floor and the ceiling, they are equidistant. They can be used to calculate the sound pressure level in flat rooms (see Chapter 95).

If perpendicular side walls are added that are also perpendicular to each other, we arrive at the rectangular room already considered in Section 2.2. Figure 10 represents one plane of the pattern of image rooms and image sources. Of course, this lattice has to be completed in the third dimension; that is, there are infinitely many patterns of this kind one upon the other. Each image of the room contains exactly one image source, and each of them is visible from all positions within the enclosure. A general property of the energy impulse response can be easily derived from this picture: All contributions arriving in the time interval between t and $t + dt$ are due to image sources located in a spherical shell with radius $R = ct$. Since the volume of a thin shell increases with the square of its radius, the temporal density of reflected components in a diagram similar to Fig. 8 grows proportionally with the square of the time delay t. This increase is balanced by the inverse-square law of spherical spreading, and the influence of wall absorption (and attenuation in the medium) leads to an exponential decrease of the energy received per second.

3.3 Enclosures with Diffusely Reflecting Walls

Very often some or all boundaries of an enclosure have a structured surface with irregularities the dimensions of which are not very small compared with the acoustic wavelength. Then at least a fraction of the incident sound energy will not be reflected specularly but will be scattered into many different directions. This process is called *diffuse reflection*. In the limiting case all the energy is diffusely reflected and no specular component will occur at all. In this case it is often convenient to apply Lambert's law, according to which the scattered intensity is proportional to the cosine of the scattering angle, and accordingly the polar diagram of the scattered intensity (see Fig. 11) is a circle (or rather a sphere). This is expressed by the following formula, in which B is the *irradiation strength* of a wall element dS, that is, the energy arriving at the wall per second and unit area. Then the scattered intensity at a point P a distance r from the wall element is

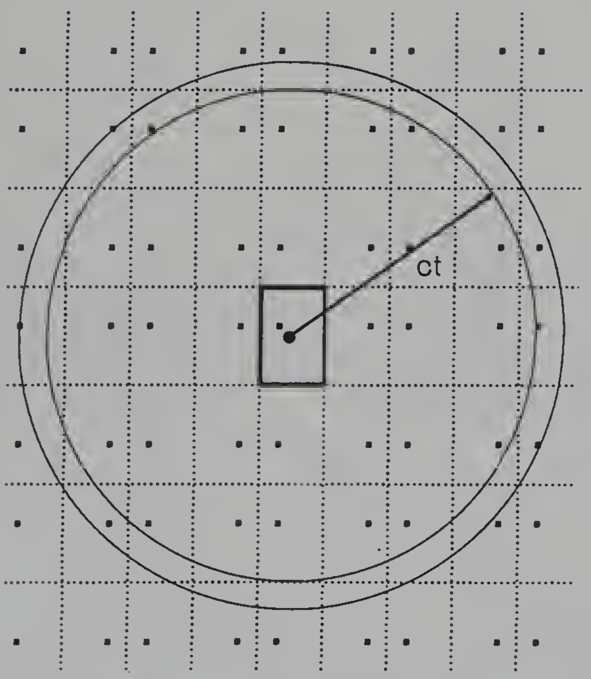


Fig. 10 Rectangular room and its image sources (one plane).

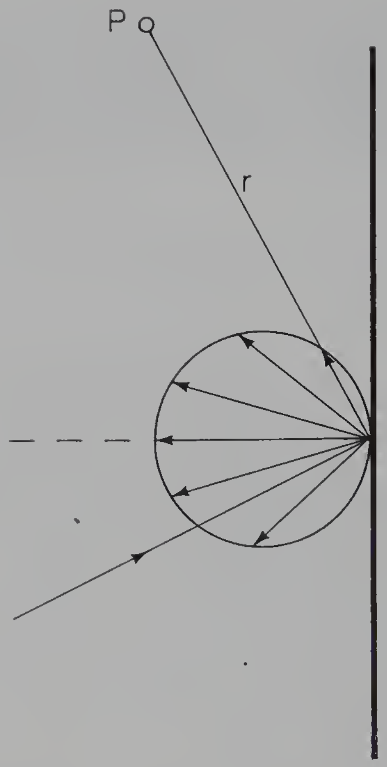


Fig. 11 Diffuse reflection according to Lambert's law.

$$dI(r, \vartheta) = B(1 - \alpha)\, dS\, \frac{\cos \vartheta}{\pi r^2}. \qquad (19)$$

As before, α denotes the absorption coefficient while ϑ is the scattering angle.

Neither the concept of sound rays nor that of image sources can be applied to diffusely reflecting boundaries because each sound ray will be split up into infinitely many secondary rays when it hits the boundary. The scattering leads to strong mixing of the sound energy, and as a result, the distribution of sound energy will be much more uniform than with specularly reflecting walls.

Let us now suppose that Lambert's law according to Eq. (19) can be applied to all wall reflections. Then the following integral equation for the *irradiation strength B* can be set up[7,8]:

$$B(\mathbf{r},t)=\iint\limits_S K(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')B\left(\mathbf{r}',t-\frac{R}{c}\right)dS' + B_0(\mathbf{r},t). \qquad (20)$$

In this equation $\mathbf{r}$ and $\mathbf{r}'$ denote the coordinates of the boundary elements dS and dS', respectively (see Fig. 12), and R is their distance. The term $B_0(\mathbf{r})$ is the "direct" irradiation strength produced by a sound source, and $\rho = 1 - \alpha$ denotes the *reflection coefficient* of the wall at $\mathbf{r}$. The kernel $K(\mathbf{r},\mathbf{r}')$ of the integral equation contains Lambert's law and also the inverse-square law of sound propagation. It is given by

$$K(\mathbf{r},\mathbf{r}')=\frac{\cos\vartheta\cos\vartheta'}{\pi R^2}, \qquad (21)$$

where ϑ and ϑ' are the angles between the boundary in dS and dS' and the straight line connecting them. The double integral in Eq. (20) is extended over the whole boundary S.

Equation (20) means simply that the sound energy arriving at a certain wall element dS at time t is the sum of the contributions that have been reemitted previously (at time $t - R/c$) by all other wall elements dS'. The effects of sound attenuation in the medium can be included by an additional factor $\exp(-mR)$ on the right side of Eq. (21), where m denotes the intensity-related attenuation constant.

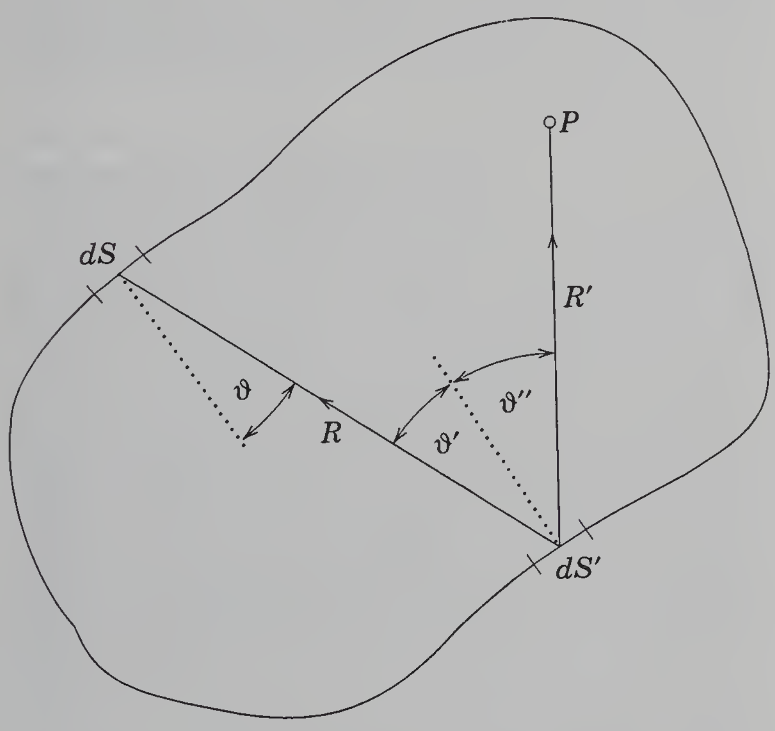


Fig. 12 Sound propagation in an enclosure with diffusely reflecting boundary: dS, receiving wall element; dS', emitting wall element; P, interior point.

If the sound power produced by the sound source is constant in time, neither B_0 nor B will depend on time; hence the steady-state version of Eq. (20) reads

$$B(\mathbf{r})=\iint\limits_S K(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')B(\mathbf{r}')\,dS' + B_0(\mathbf{r}). \qquad (22)$$

On the other hand, for $B_0 = 0$, Eq. (20) describes the sound decay in the enclosure. If the irradiation strength B on the boundary is known, the energy density in any point $\mathbf{r}''$ within the enclosure can be calculated[9]:

$$u(\mathbf{r}'',t)=\frac{1}{\pi c}\iint\limits_S \rho(\mathbf{r}')B\left(\mathbf{r}',t-\frac{R'}{c}\right)\frac{\cos\vartheta''}{R'^2}\,dS' + u_0(\mathbf{r}'',t), \qquad (23)$$

where ϑ'' denotes the angle between the wall normal and the line with length $R' = |\mathbf{r}'' - \mathbf{r}'|$ that connects the wall element dS' with the inner point. The function u_0 is the energy density due to the direct sound. For steady-state conditions, the time dependence of B is just neglected. A similar expression holds for the sound intensity.

Generally, Eq. (20) or (22) must be numerically solved, which is not very difficult since standard procedures are available. The steady-state case has been discussed by Carroll and Miles,[10] who presented a general, yet formal solution to Eq. (22). Numerical calculations for the rectangular room have been carried out by Miles,[11] who treated the sound distribution under steady-state conditions as well as in decaying sound fields. Closed solutions can be worked out only for simple room shapes such as spherical enclosures.[8] Another example for which a closed solution can be obtained, at least for stationary sound propagation, is for the infinite flat room[12] already considered in Section 3.2. But now both walls or one of them is assumed as diffusely reflecting, which yields a more realistic model of a real hall than that with smooth walls since in factories or bureaus there are machines, piles of materials, furniture, and so on, which cause noticeable scattering of sound energy. For this flat room, the kernel $K(\mathbf{r},\mathbf{r}')$ in Eq. (22) is just $a^2/\pi R^4$. More about the solution may be found in Chapter 95.

4 STATISTICAL ACOUSTICS

In this section closed formulas for the energy density in an enclosure both at steady-state conditions and during sound decay are presented. Such expressions are of considerable importance in practical room acoustics, for example, to predict noise levels in working environments or to assess the suitability of a room for certain types of performances.

As in the preceding section, only energy propagation will be discussed here; that is, no phase relations between sound waves will be taken into account. Therefore the sound field components we consider are sound rays; alternatively, even the somewhat hypothetical notion of "sound particles" will be used. (These should not be identified with "phonons"; their physical meaning is rather that of short sound impulses or bursts with wide bandwidth traveling along a sound ray.) However, in contrast to the discussions of the preceding section, energy densities are not obtained by adding the energies of individual sound rays or sound particles but by considering their "average fate."

Most of the laws to be discussed in this section are valid under the assumption of a homogeneous or diffuse sound field, a condition that is only incompletely fulfilled in practical situations. Therefore, the formulas and laws derived under this assumption cannot be expected to be quite accurate. In fact, deviations from complete diffusivity influence the reliability of the expressions differently, depending on whether we regard steady-state parameters or reverberation time; in the latter case the requirements concerning the diffusion or diffusivity of the sound field are less stringent.

Because of the central importance of the diffuse sound field, in the following some basic properties of diffuse sound fields will be presented.

4.1 Basic Properties of Diffuse Sound Field

We speak of a perfectly diffuse sound field if the sound energy arriving at any point inside the enclosure is uniformly distributed over all possible directions of incidence. In this case the intensity vector in the field is zero everywhere. It is clear that this is an ideal condition that cannot be exactly fulfilled because in a real enclosure there are losses in the medium and at the boundary therefore at least some energy transport from the sound source into the medium and toward the walls must occur.

It is easy to show that, under the assumption of a diffuse field, the energy density is locally constant. Furthermore, each point of the boundary is hit by the same amount of energy per second and unit area, and if we call the latter quantity irradiation strength and denote it by B as in Section 3.3, it can be shown that

$$B = \frac{c}{4}\, u. \tag{24}$$

From this relation the average number $\overline{n}$ of boundary reflections a sound ray or a sound particle undergoes per second can be obtained:

$$\overline{n} = \frac{cS}{4V}, \tag{25}$$

where V is, as before, the volume of the enclosure and S is the area of its boundary. Then, the average length of a sound ray between two reflections from the boundary, or the mean free path length of a sound particle, is

$$\overline{l} = \frac{4V}{S}. \tag{26}$$

At each boundary point the sound energy arrives uniformly from all directions of the half space. Therefore the effective absorption coefficient α_d is obtained by suitably averaging the angle-dependent absorption coefficient $\alpha(\vartheta)$:

$$\alpha_d = \int_0^{\pi/2} \alpha(\vartheta)\, \sin(2\vartheta)\, d\vartheta. \tag{27}$$

This relation is known as *Paris's formula*. For locally reacting boundaries, that is, if the wall impedance Z is independent of the direction of sound incidence, the result of this averaging can be expressed in a closed formula. Its content is shown in Fig. 13, which represents the contours of constant absorption coefficient α_d in the plane formed by the magnitude and the phase angle of the specific wall impedance ϑ. It is noteworthy that α_d will never reach unity but has an absolute maximum 0.951 for the real impedance $\zeta = 1.567$.

Finally a remark may be in place on the conditions under which a diffuse sound field can be expected. The degree of sound field diffusion depends not only on the shape of the enclosure but also on the surface structure of its boundary and the magnitude and the spatial distribution of the wall absorption. In an enclosure with the shape of an irregular polyhedron the sound rays will be mixed up to some extent and thus establish a sound field with at least some diffusion. On the contrary, we can hardly expect a diffuse sound field within a sphere with a specularly reflecting boundary. The situation will be changed, however, if the sphere has a diffusely reflecting boundary, which will equalize the energy distribution. Generally, diffuse wall reflections increase the diffusion of the sound field in the enclosure. (A similar effect will have sound-scattering objects within the enclosure.) But even then, the absorption distribution along the walls influ-

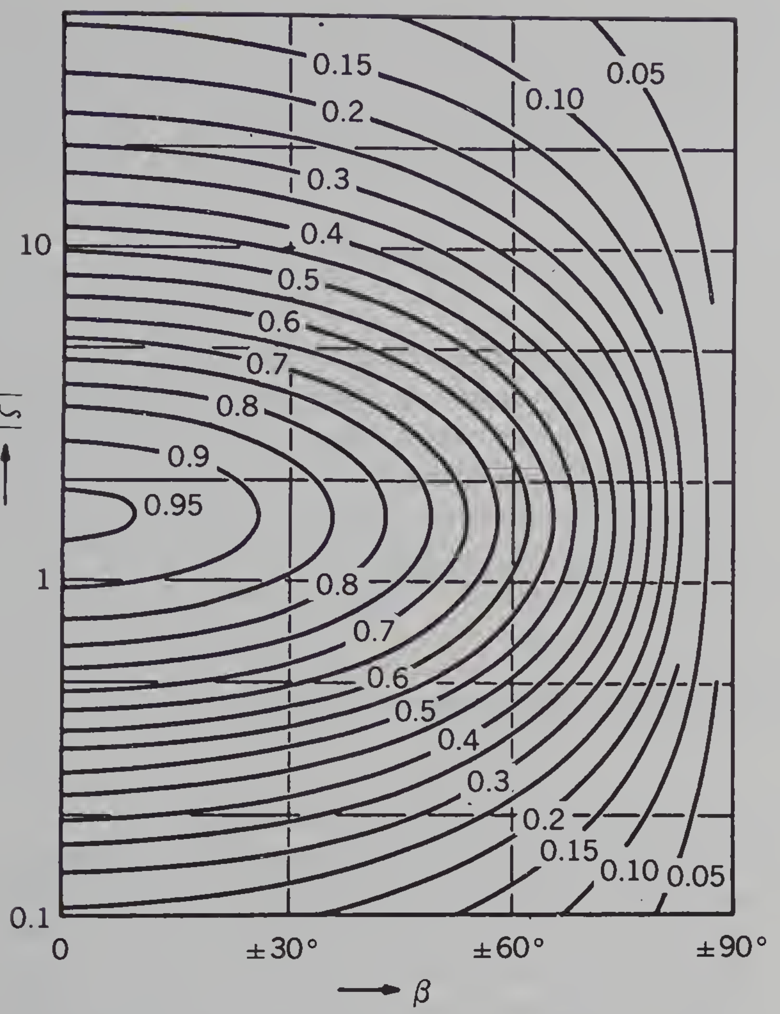


Fig. 13 Curves of equal absorption coefficient of a locally reacting wall for random sound incidence. The abscissa and the ordinate are the phase angle and the magnitude, respectively, of the specific wall impedance $\zeta = Z/\rho_0 c$.

ences the degree of diffusion. Therefore, the condition of diffuse reflections must be distinguished from that of a diffuse sound field.

4.2 Stationary Energy Density

We imagine an enclosure in which a sound source with arbitrary directivity emits the sound power P. It is assumed that the produced sound field is a diffuse one. Therefore, all absorption coefficients are those for random incidence, averaged according to Eq. (27). (For the sake of simplicity, the index d is omitted in the following.) Now suppose that the boundary can be subdivided into portions with areas S_i and with absorption coefficients α_i. Each of these portions absorbs the energy $\alpha_i S_i B = \alpha_i S_i cu/4$ per second. Under these assumptions, the temporal change of the energy density within the enclosure is

$$\frac{du}{dt} = \frac{1}{V}\left[P(t) - \frac{c}{4}Au\right]. \tag{28}$$

In this expression the quantity

$$A = \sum_i \alpha_i S_i \tag{29}$$

has been introduced, the so-called equivalent absorption area or total absorption of the enclosure. It has the dimension of square metres and can be imagined as the area of a totally absorbing boundary portion, for instance, an open window, in which the total absorption of the enclosure is concentrated. Absorption coefficients of typical wall materials may be found in Chapter 92.

Since we discuss the steady-state energy density u_s in this section, we assume constant power output P; accordingly the time derivative of u will vanish. Then Eq. (28) reduces to

$$u_s = \frac{4P}{cA}. \tag{30}$$

This relation is the basis of an important method to determine experimentally the total power output P of a sound source by measuring the energy density. For this purpose it may be convenient to convert the above relation into the logarithmic scale and thus to connect the sound pressure level SPL with the power level PL $= 10\log_{10}(P/10^{-12}\text{ W})$:

$$\text{SPL} = \text{PL} - 10\log_{10}\left(\frac{A}{1\text{m}^2}\right) + 6\text{ dB}. \tag{31}$$

The equivalent absorption area A of the enclosure used for this procedure is obtained by measuring its reverberation time (see next section).

Equations (30) and (31) are valid as long as the distance r of the observation point from the sound source is relatively large. In the vicinity of the sound source, however, the contribution of the direct component to the energy density is predominant. For a nondirectional sound source this is

$$u_d = \frac{P}{4\pi c r^2}. \tag{32}$$

In a certain distance, the so-called reverberation distance r_r, both components, namely the direct and the reverberant parts of the energy density, are equal (see Fig. 14):

$$r_r = \frac{1}{4}\sqrt{\frac{A}{\pi}} \approx 0.057\sqrt{\frac{V}{T}}. \tag{33}$$

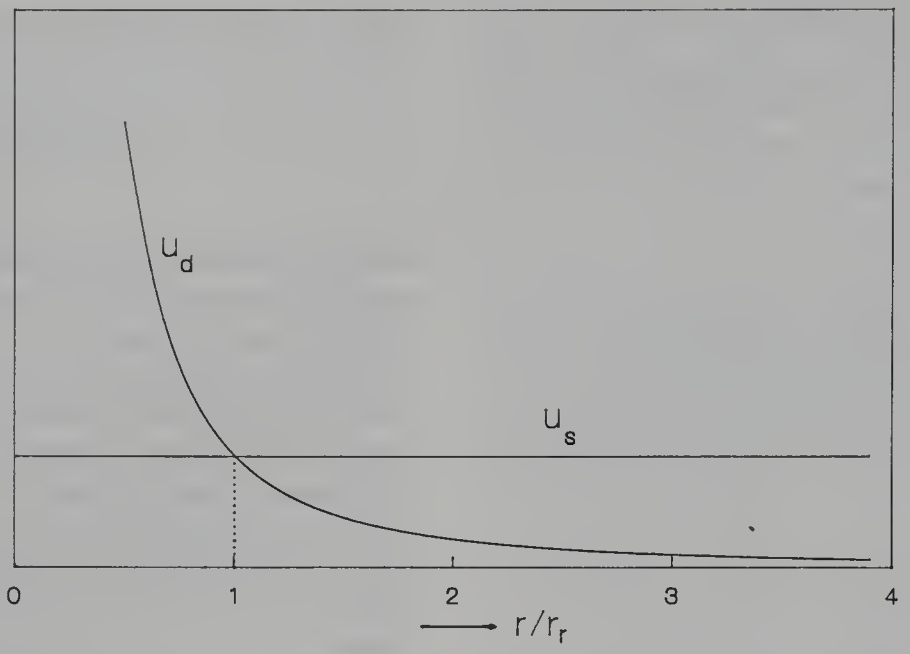


Fig. 14 Direct energy density u_d and reverberant energy density u_s in an enclosure as a function of the ratio r/r_r.

Hence, a somewhat more general expression for the total energy density including the direct component is

$$u = \frac{P}{4\pi c}\left(\frac{1}{r^2} + \frac{1}{r_r^2}\right). \tag{34}$$

If the sound source has a certain directivity, the first term in the parentheses has to be multiplied with its *directivity factor* Γ, which compares the intensity radiated into the direction of observation with the average intensity, both for the same distance. Hence, the corresponding expression for the sound pressure level reads

$$\mathrm{SPL} = \mathrm{PL} - 20\log_{10}\left(\frac{r}{1\mathrm{m}}\right) + 10\log_{10}\left(\Gamma + \frac{r^2}{r_r^2}\right) - 11\ \mathrm{dB}. \tag{35}$$

The above relations should be applied with some caution. For sine signals or other signals with small-frequency bandwidth they yield an average, at best, since we know from the discussions in Section 2.3 that the sound pressure amplitudes as well as the energy densities in an enclosure are far from being constant but are distributed over a wide range. But even for wide-band excitation of the enclosure, the energy density is usually not constant throughout the enclosure. One particular deviation known as the *Waterhouse effect* is due to the fact that any reflecting wall enforces certain phase relations between incident and reflected waves no matter from which directions the waves arrive at the wall.[13] As a consequence, there are fluctuations of the energy density in the vicinity of a wall that depend on its acoustical properties and on the frequency spectrum of the sound signal. In any case, the energy density in front of a rigid wall is twice its average value far away from it. Since the range of these fluctuations is about one-quarter of a wavelength, they can be neglected in the high-frequency limit.

A relatively high degree of sound field diffusion can be achieved in reverberation chambers with highly reflective walls and with an irregular shape that supports the mixing and blending of sound. For such a room Eq. (31) may be applied with some precautions. In other cases, it is preferable to compute the steady-state energy density by applying the image source model described in Section 3 if the walls are specularly reflecting or by numerically solving Eq. (22) and inserting the result into Eq. (23), provided the boundary reflection is at least partially diffuse.

4.3 Sound Decay

Now we consider decaying sound fields. For our discussion we imagine that a sound source exciting the enclosure until time $t = 0$ is abruptly stopped. (As an alternative, such a decaying sound field can be produced by a short impulse emitted at $t = 0$.) If the sound field can be expected to be diffuse, Eq. (28) with $P = 0$ can be applied, which produces the energy density

$$u(t) = u_0 \exp\left(-\frac{cA}{4V}t\right) \quad \text{for } t \geq 0, \tag{36}$$

where u_0 is the initial energy density at $t = 0$. From this equation, the reverberation time T of the enclosure, that is, the time in which the energy density has dropped by a factor of 10^6 (see Section 2.4), is obtained:

$$T = 0.163 \frac{V}{A}. \tag{37}$$

The numerical constant in this formula is 24 ln $10/c$ and has the dimension seconds per metre. Accordingly, all lengths have to be expressed in metres.

Equation (37) is the famous Sabine reverberation formula and is probably the most important relation in room acoustics. Despite its simplicity, its accuracy is sufficient for many practical applications provided the correct value for the equivalent absorption A is inserted. However, for enclosures with relatively high absorption, it fails: For an absorption coefficient 1 on all walls, it predicts a finite reverberation time, although it is clear that in this case there will be no reverberation at all.

Therefore more accurate reverberation formulas have been derived, taking into account that sound decay is not a continuous process as assumed in Eq. (28) but involves stepwise reduction of a sound particle's energy whenever it is reflected from a wall.[14] Let t be the time that has elapsed since the beginning of the decay process. Then a typical sound particle will have undergone $\overline{n}t$ reflections from the boundary, with $\overline{n}$ denoting the average number of wall collisions per second, according to Eq. (25), and its energy will have diminished by the factor $(1 - \alpha)^{\overline{n}t}$. Because of the assumed diffusivity of the sound field, this particle can be regarded representative for all sound particles. Therefore the same reduction factor holds for the whole energy within the enclosure:

$$u(t) = u_0(1 - \alpha)^{\overline{n}t} \quad \text{for } t \geq 0. \tag{38}$$

For nonuniform distribution of the absorption along the boundary, the absorption coefficient α is replaced with its arithmetic average:

$$\overline{\alpha} = \frac{A}{S} = \frac{1}{S} \sum_i \alpha_i S_i. \tag{39}$$

The reverberation time obtained from this relation is

$$T = 0.163 \frac{V}{-S \ln(1 - \overline{\alpha})}. \tag{40}$$

This formula is usually known as Eyring's formula, although it was described earlier by Fokker (1924) and by Waetzmann and Schuster (1929). It is obvious that it yields the reverberation time $T = 0$ for a totally absorbing boundary. For $\overline{\alpha} \ll 1$, it is identical to Sabine's formula [Eq. (37)].

Sound attenuation in the free medium can be taken into account by an additional factor $\exp(-mct)$ in Eq. (36). This modifies the above formula for the reverberation time in the following way:

$$T = 0.163 \frac{V}{4mV - S \ln(1 - \overline{\alpha})}. \tag{41}$$

In Sabine's formula, Eq. (37), the attenuation by the medium can also be included by adding the term $4mV$ in the denominator.

Several authors[15,16] proposed not to average the absorption coefficients arithmetically, as in Eq. (39), but to average the logarithm of the *reflection coefficients* $1 - \alpha_i$. This leads to

$$T = 0.163 \frac{V}{-\sum_i S_i \ln(1 - \alpha_i)}. \tag{42}$$

This is the reverberation formula of Millington and Sette. It has the strange consequence that the reverberation time will be zero if any part of the boundary has the absorption coefficient 1. A more rigorous derivation shows, however, that the averaging procedure prescribed by Eq. (39) of Chapter 77 is the correct one.

It should be recalled that the derivation of all the formulas presented in this section was based upon the assumption of a diffuse sound field that ensures that each element of the enclosure is hit by the same energy amount per second. On the other hand, it was pointed out in Section 4.1 that sound fields in real rooms will fulfill this condition at best approximately because of the inevitable wall and medium losses. Even stronger violations are brought about by particular room shapes, such as those of very long or very flat rooms, or by a highly nonuniform distribution of the absorption along the boundary. In an occupied hall, for instance, the absorption is concentrated onto the area where the audience is seated. In such cases we must expect that the application of one of the reverberation formulas above would lead to unpredictable errors.

If the assumption of a diffuse sound field is dropped, the energy decay could be computed by evaluating Eq. (20), which reduces to a simple recursion if it is expressed in discrete form. In any case, the final sound decay follows an exponential law with the same decay rate at each location, as has been shown by Miles.[11] Several effective iteration schemes have been developed on the basis of Eq. (20)[9,17] to determine this decay rate and hence the reverberation time of the enclosure. Strictly speaking, however, Eq. (20) holds only for enclosures with diffusely reflecting boundaries. Nevertheless, the mentioned iteration schemes can be applied as well to practical situations where wall reflections are only partially diffuse. In this case the condition of diffuse reflections is reached in several steps, so to speak, since the conversion of specularly reflected into diffusely reflected sound is irreversible. Moreover, it has been shown by Hogson[18] that in nearly all halls there is at least some diffuse wall reflection.

To summarize the discussions of this section, it can be stated that the reverberation time of an enclosure can be predicted or calculated quite well. In many cases it is sufficient to apply one of the simple formulas for the decay time. But even if this is not the case, more reliable methods can be applied with still reasonable computing effort. Concerning the steady-state sound level or energy density, the situation is less satisfactory, since the reliability of simple formulas such as Eq. (34) depends more critically on the diffusion of the sound field than the relations for the decay time. Therefore, for computation of the steady-state level, often more complicated schemes are needed, such as the image source model.

REFERENCES

1. J. Pan and D. A. Bies, "The Effect of Fluid-Structural Coupling on Sound Waves in an Enclosure," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 691–707 (theoretical part), pp. 708–717 (experimental part).
2. M. R. Schroeder and H. Kuttruff, "On Frequency Response Curves in Rooms. Comparison of Experimental, Theoretical and Monte Carlo Results for the Average Spacing between Maxima," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 76–80.
3. H. Kuttruff and R. Thiele, "Über die Frequenzabhängigkeit des Schalldrucks in Räumen," *Acustica*, Vol. 4 (Beihefte), 1954, pp. 614–617.
4. M. Schröder, "Die statistischen Parameter der Frequenzkurven von großen Räumen," *Acustica*, Vol. 4 (Beihefte), 1954, pp. 594–600.
5. J. Borish, "Extension of the Image Source Model to Arbitrary Polyhedra," *J. Acoust. Soc. Am.*, Vol. 75, 1978, pp. 1827–1836.
6. H. Lehnert, *Binaurale Raumsimulation: Ein Computermodell zur Erzeugung virtueller auditiver Umgebungen*, Shaker, Aachen, 1992.
7. H. Kuttruff, "Simulierte Nachhallkurven in Rechteckräumen," *Acustica*, Vol. 25, 1971, pp. 333–342; "Nachhall und effektive Absorption in Räumen mit diffuser Wandreflexion," *Acustica*, Vol. 35, 1976, pp. 141–153.
8. W. B. Joyce, "Effect of Surface Roughness on the Reverberation Time of a Uniformly Absorbing Spherical Enclosure," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 1429–1436.
9. H. Kuttruff, "A Simple Iteration Scheme for the Computation of Decay Constants in Enclosures with Diffusely Reflecting Boundaries," *J. Acoust. Soc. Am.*, Vol. 98, 1995, pp. 288–293.
10. M. M. Carroll and R. N. Miles, "Steady-State Sound in an Enclosure with Diffusely Reflecting Boundary," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 424–1428.
11. R. N. Miles, "Sound Field in a Rectangular Enclosure with Diffusely Reflecting Boundaries," *J. Sound Vib.*, Vol. 92, 1984, pp. 203–213.
12. H. Kuttruff, "Stationäre Schallausbreitung in Flachräumen," *Acustica*, Vol. 57, 1985, pp. 62–70.
13. R. V. Waterhouse, "Interference Patterns in Reverberant Sound Fields," *J. Acoust. Soc. Am.*, Vol. 27, 1955, pp. 247–258.
14. C. F. Eyring, "Reverberation Time in 'Dead' Rooms," *J. Acoust. Soc. Am.*, Vol. 1, 1930, pp. 217–241; "Methods of Calculating the Average Coefficient of Sound Absorption," *J. Acoust. Soc. Am.*, Vol. 4, 1993, pp. 178–192.
15. G. Millington, "A Modified Formula for Reverberation," *J. Acoust. Soc. Am.*, Vol. 4, 1932, pp. 69–82.
16. W. J. Sette, "A New Reverberation Time Formula," *J. Acoust. Soc. Am.*, Vol. 4, 1933, pp. 193–210.
17. E. N. Gilbert, "An Iterative Calculation of Reverberation Time," *J. Acoust. Soc. Am.*, Vol. 69, 1981, pp. 178–184.
18. M. R. Hodgson, "Evidence of Diffuse Surface Reflection in Rooms," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 765–771.

92

SOUND ABSORPTION IN ENCLOSURES

DAVID A. BIES AND COLIN H. HANSEN

1 INTRODUCTION

Sound absorption in enclosures occurs when sound waves strike objects in the enclosure and the enclosure boundaries as well as during propagation through the acoustic medium (usually air) that fills the enclosing space.

Sound absorption in enclosures plays an important part in the determination of sound pressure levels resulting from the operation of sound sources of known sound power output as well as in determining the amount of reverberation, or "liveliness," of the enclosure, which is quantified in terms of its reverberation time. Thus this chapter will be concerned with the measurement and calculation of sound absorption in enclosures as well as the estimation of the effect of sound absorption on sound levels and reverberation times. Relationships between sound absorption, acoustic impedance, and flow resistance for acoustically absorptive materials will also be discussed.

The physical mechanism associated with sound absorption at the enclosure boundaries may take one of two forms, dependent upon the vibroacoustic properties of the room boundaries, the frequency range of interest, and the size of the enclosure. One of the mechanisms is referred to as *local reaction*, meaning that the response of the boundary to an incident sound wave (and thus the absorption properties) is entirely dependent upon the local properties of the boundary at the point of incidence. In this case the boundary absorption may be described in terms of a sound absorption coefficient that is the ratio of energy absorbed to the energy incident.

The second mechanism of absorption that can occur when a sound wave is incident on a boundary is one of modal coupling, which means that the sound field in the enclosure excites the boundaries into vibration, which in turn results in removal of energy from the sound field and effective absorption by the boundaries. Whether the absorption mechanism is of the first or second kind will determine the type of analysis needed to evaluate its effect on the enclosed sound field. Conditions that need to be satisfied for each of the absorption mechanisms to occur will be discussed in the following paragraphs.

2 MODAL PROPERTIES OF ENCLOSURES

Sound absorption in enclosures is dependent upon the modal response of the enclosed space, and thus what is meant by a *mode* is important to the understanding of sound absorption. Modal response in enclosures is discussed in Chapter 91. As shown in the latter chapter, the case of a rectangular enclosure allows an exact solution of the wave equation in terms of the enclosure dimensions. It should be noted that the conclusions drawn from consideration of a rectangular enclosure are, however, by no means restricted to rectangular enclosures but apply to enclosures of any shape. The analysis leads to a simple expression for calculating the frequencies of the normal modes of a rectangular enclosure [see Chapter 91, Eq. (6)] as well as the total number of modes within a frequency band [see Chapter 91, Eq. (7)]. It may be shown that the number of modes N resonant up to and including frequency f in a room of any shape is a function only of the room volume V, the room surface area S, the room perimeter P (when properly defined), and speed of sound c. Thus the modal density, defined as the number of modes per unit frequency, $\partial N/\partial f$, in a room

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

of any shape is given by[1]

$$n_A = \frac{4\pi V}{c^3} f^2 + \frac{\pi S}{2c^2} f + \frac{P}{8c}. \qquad (1)$$

The absorption mechanism involving enclosure–boundary modal coupling is important when the enclosure modal density is low and wall damping is also small. This mechanism thus characterizes vehicles of various kinds (e.g., aircraft and automobiles); it is not amenable to generalization, has received only limited investigation, and is beyond the scope of this chapter, except for the discussion of panel absorbers at the end.

If the enclosure boundaries are well damped or if the modal density is sufficiently high that the sound field may be regarded as diffuse, then the absorption mechanism will be dominated by the local reaction effect described earlier. The modal density may be regarded as sufficiently high when the quadratic term in frequency of Eq. (1), which describes the modal density of oblique modes, is larger than the linear and constant terms, indicating that oblique modes dominate the enclosure response.

A more precise definition of the conditions under which the sound field may be considered diffuse is related to damping of the modes in the room as well as to the modal density. The modal damping can be quantified in terms of the modal bandwidth Δf, as shown in Fig. 1. As can be seen from the figure, the modal bandwidth for a particular mode is the width in hertz of the response curve of the mode measured 3 dB down from the peak response. The modal bandwidth (averaged over all modes in a frequency band B) is defined in terms of the room reverberation time T_{60} for the frequency band B as follows[2]:

$$\Delta f = \frac{2.2}{T_{60}}, \qquad (2)$$

where the reverberation time is defined as the time taken for the sound field to decay by 60 dB after the sound source is shut off.

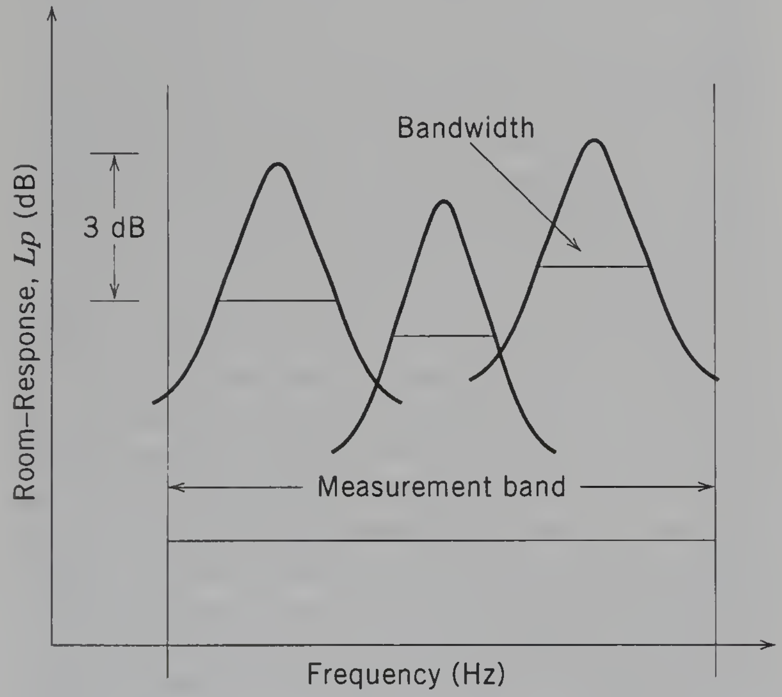


Fig. 1 Illustration of modal overlap.

Referring to Fig. 1, the modal overlap (or the extent to which the modes result in a uniform pressure distribution as a function of frequency) is defined as

$$M = \frac{\Delta f_1 + \Delta f_2 + \Delta f_3}{B}, \qquad (3)$$

or in more general terms as[3]

$$M = \Delta f n_A. \qquad (4)$$

For the purposes of the present discussion, a value of modal overlap greater than 3 will mean that the sound field is sufficiently diffuse that the local reaction mechanism of sound absorption will dominate. The frequency range for which this occurs will be referred to as the high-frequency range. It is of interest to note that most auditoria and industrial spaces are sufficiently large and sufficiently well damped that the entire audio range generally of interest (63 Hz to 8000 Hz) may be considered as the high-frequency range.

The discussion to follow will be restricted to enclosures for which no dimension exceeds the others by more than about a factor of 3. Such enclosures will be described here as Sabine enclosures; enclosures that do not satisfy this restriction are discussed in Chapter 95. It is important to note that the Sabine room can be characterized by a reverberant sound pressure level L_{pR} and a reverberation time T_{60} and as such provides a means for the definition, measurement, and use of the Sabine absorption coefficient of materials.

3 RELATIONSHIP BETWEEN SOUND ABSORPTION AND ENCLOSURE SOUND PRESSURE LEVEL

In the high-frequency range, the sound field in a Sabine room may be though of as consisting of two fields superimposed upon each other. These fields are, respectively, the direct field of the source and the reverberant field of the room. The latter field is the result of multireflections at the boundaries of the room and is in an average sense uniform throughout. The direct field of a source is greatest near the source and dominates in some area about the source, but as it spreads out into the room, it necessarily decays until it in turn is dominated by the reverberant field in the room.

The simplification will be made that a source may be represented as a point source multiplied by a directivity factor D_i and a sound power weighting factor d_i. When account is taken of n sources, an expression for the sound pressure level L_p at any point in a Sabine room in terms of the total sound power level L_w of all sources is as follows[3]:

$$L_p = L_w + 10\ \log_{10}\left[\sum_{i=1}^{n} \frac{d_i D_i}{4\pi r_i^2} + \frac{4}{R}\right]. \tag{5}$$

In Eq. (5), the weighting factor d_i and the directivity factor D_i for the ith source have been introduced. The weighting factor is calculated as $d_i = W_i/W$, where W_i is the sound power of the ith source and W is the arithmetic sum of all sources, both in consistent units. The directivity factor D_i may be taken as 1 for a source raised up off the floor and 2 for a source on the floor. Alternatively, if the directional properties of the source are known, then this information may be used instead. The quantities r_i are the linear distances from the assumed acoustical centers of the several sources to the point of observation.

The room constant R introduced in Eq. (5) is calculated in terms of the mean Sabine absorption coefficient $\overline{\alpha}$ and the total area S of all absorbing surfaces S_i as follows.[3] The determination of $\overline{\alpha}$ is discussed in Section 5:

$$R = \frac{S\overline{\alpha}}{1 - \overline{\alpha}}, \tag{6}$$

where values of Sabine absorption coefficients $\overline{\alpha}_i$ for areas S_i are known, perhaps by reference to a table of measured values, and the mean Sabine absorption coefficient $\overline{\alpha}$ is calculated as the area-weighted average as

$$\overline{\alpha} = \sum_{i=1}^{n} \frac{S_i}{S}\ \overline{\alpha}_i. \tag{7}$$

The quantity $S_i\alpha_i$ has the units of square feet or square metres, but in either case it has the name *sabin*. Here it will be called a metric sabin when it has the units of square metres.

The direct field has been described as the field close to a source where it dominates the reverberant field. Close to the ith source, the distance r_i will usually be small compared to the distances r_j of all other sources. Consequently, the ith term will usually be larger than all other terms in the sum [see Eq. (7)] and also larger than the reverberant field term. Close to the ith source, r_i satisfies the requirement

$$\frac{1}{4}\sqrt{\frac{Rd_iD_i}{\pi}} \gg r_i. \tag{8}$$

In the range satisfied by Eq. (8), the sound field is controlled by the ith source, and consequently any absorptive treatment that decreases the reverberant field will have an insignificant effect upon the observed sound level in this range. For example, if an employee works close to a machine in its direct field, the placement of absorptive treatment upon the walls or ceiling of the room will have no significant effect upon the noise to which the employee is exposed. Alternatively, if the employee's machine is close to a reflecting wall, treatment of the wall will reduce the reflected sound from the wall, but this could at most decrease the employee's exposure on a sound energy basis by half, resulting in a 3-dB sound reduction. This reduction would barely be noticeable.

The reverberant field exists well removed from all sources. For example, when the point of observation is closest to the ith source and in the reverberant field, the criterion that is met is as follows:

$$r_i \gg \frac{1}{4}\sqrt{\frac{Rd_iD_i}{\pi}}. \tag{9}$$

In the reverberant field that satisfies Eq. (9), the sound pressure level is controlled by the room constant R, which in turn is controlled by the mean Sabine absorption coefficient. The placement of sound absorbents on the surfaces of the room will reduce the sound in the reverberant field to the extent that the mean Sabine absorption coefficient is reduced. Reference to Eq. (7) shows that the mean Sabine absorption coefficient is an area-weighted quantity. Consequently, best results are obtained when least absorptive surfaces are treated with absorbent material.

The reduction in sound pressure level L_p in the reverberant field due to treatment with sound-absorbing material may be calculated using the following equation[1]:

$$\Delta L_p = 10\ \log_{10} \frac{R_f}{R_i}. \tag{10}$$

Here R_i is the initial value and R_f is the final value of the room constant after the addition of absorbent.

4 EFFECT OF SOUND ABSORPTION ON REVERBERATION TIME

When sound is introduced into a room, the reverberant field level will increase until the sound energy introduction is just equal to the sound energy absorption. If the sound source is abruptly shut off, the reverberant field will decay at a rate determined by the rate of sound energy absorption. The time required for the reverberant field to decay by 60 dB is called the reverberation time and is the single most important parameter describing the acoustical properties of a room.

These observations may be summarized as follows, where it will be assumed that sound is absorbed at the boundaries of the room. The power W_a absorbed by the room boundaries is equal to the product of the boundary area S, the mean Sabine absorption coefficient $\overline{\alpha}$, and the amplitude of the sound intensity component normally incident at the boundary surface. Thus,[3]

$$W_a = IS\overline{\alpha} = \frac{\psi S c\overline{\alpha}}{4}. \tag{11}$$

Note that although the intensity averaged over all directions (the net intensity) is zero[1], the acoustic intensity component in a particular direction in a diffuse field is nonzero and as given above.

The introduced sound power W less the sound power absorbed is equal to the rate of change of stored sound energy in the room. Thus,

$$W - \frac{\psi S c\overline{\alpha}}{4} = V\frac{\partial\psi}{\partial t}. \tag{12}$$

In Eq. (12), $\psi(t)$ is the sound energy density at time t, S is the surface area of the walls of the room, c is the speed of sound, V is the volume of the room, and $\overline{\alpha}$ (the fraction of incident energy absorbed) is the mean sound absorption coefficient for the walls of the room. The sound energy density is related to the mean-square sound pressure $\langle p^2\rangle$, gas density ρ, and speed of sound c and is a function of time t as follows:

$$\psi(t) = \frac{\langle p^2(t)\rangle}{\rho c^2}. \tag{13}$$

At time zero, when equilibrium exists between power input and power absorbed, the source is abruptly shut off, so that $W = 0$, and the sound field decays. The mean-square sound pressure at time t at any point in the room is then[1]

$$\langle p^2(t)\rangle = \langle p^2(0)\rangle e^{-Sc\overline{\alpha}t/4V}. \tag{14}$$

The time required for the sound field to decay by 60 dB, called the reverberation time T_{60}, may be determined by solving Eq. (14). The resulting equation, originally derived by Sabine,[4] gives the relation between the reverberation time and the mean sound absorption coefficient. The latter quantity is given the name *Sabine absorption coefficient*[4]:

$$T_{60} = \frac{55.25V}{Sc\overline{\alpha}}. \tag{15}$$

To gain some insight into the assumptions implicit in the derivation of Eq. (15), it is useful to repeat the analysis using a modal description of the reverberant field within an enclosure.[5] In fact, if any enclosure is driven at a frequency slightly off resonance and the source is abruptly shut off, the frequency of the decaying field will be observed to shift to the resonance frequency of the driven mode. The modal analysis requires the introduction of the modal mean free path L_i and the modal mean sound reflection coefficient β_i. The modal mean free path L_i is the mean distance between reflections of a sound wave traveling around a closed modal circuit. The modal mean reflection coefficient is the product of the reflection coefficients encountered by a wave traveling around a modal circuit.

It may be shown that the mean of the modal mean free paths is given by the equation[3]

$$\frac{1}{N}\sum_{i=1}^{N} L_i = \frac{4V}{S}. \tag{16}$$

The assumption will be made that on average, at zero time, the energy distribution among the modes of the room will be about equal; thus,

$$\langle p_i(0)^2\rangle = \frac{\langle p(0)^2\rangle}{N}. \tag{17}$$

The following equation may be shown to describe a modally decaying sound field:

$$\langle p(t)^2\rangle = \langle p(0)^2\rangle \frac{1}{N}\sum_{i=1}^{N} \exp\left(\frac{ct}{L_i}\log_e \beta_i\right). \tag{18}$$

Some simplifications will now be made. In Eq. (18), all L_i will be replaced with the mean free path $4V/S$ [see Eq. (16)] and all modal reflection coefficients β_i will be replaced with $1 - \alpha_{\text{st}}$, where α_{st} is the statistical absorption coefficient. Equation (18) may be rewritten as

$$\langle p(t)^2 \rangle = \langle p(0)^2 \rangle \exp\left(\frac{Sct}{4V} \log_e(1 - \alpha_{\text{st}}) \right). \qquad (19)$$

Equation (19) may be solved for the reverberation time and the equation of Norris Eyring is obtained as

$$T_{60} = \frac{55.25V}{-Sc \log_e(1 - \alpha_{\text{st}})}. \qquad (20)$$

The statistical absorption coefficient α_{st} for the entire enclosure is an area-weighted quantity. Where α_{st_i} is the statistical coefficient of subarea S_i, and the total room area S is the sum of all n subareas, the statistical absorption coefficient for the entire room is determined as

$$\alpha_{\text{st}} = \sum_{i=1}^{n} \frac{S_i \alpha_{\text{st}_i}}{S}. \qquad (21)$$

When $\alpha_{\text{st}} < 0.4$, an error of less than 0.5 dB is made in the determination of the sound pressure level in Eq. (19) if the following approximation is made:

$$\alpha_{\text{st}} = -\log_e(1 - \alpha_{\text{st}}). \qquad (22)$$

Use of Eq. (22) allows Eq. (20) to be rewritten in the form of Eq. (15), suggesting that the Sabine absorption coefficient is the same as the statistical absorption coefficient. In fact, when edge diffraction is taken into account, the suggestion appears to be true.[5]

Alternatively, if the mean free path replacement is made in Eq. (18), but all of the modal mean reflection coefficients β_i are replaced with the mean statistical reflection coefficient β_{st}, then the equation of Millington[7] and Sette[8] is obtained:

$$T_{60} = \frac{55.25V}{-Sc \log_e \beta_{\text{st}}}. \qquad (23)$$

The statistical reflection coefficient is defined as

$$\beta_{\text{st}} = \left(\prod_{i=1}^{n} \beta_i \right)^{S_i/S}. \qquad (24)$$

Experience has shown that Eq. (23) provides more accurate results than Eq. (15) or (20) for predicting reverberation times for large auditoria. Note that the statistical sound absorption coefficient α_{st} must be calculated from measurements of the material flow resistance or normal impedance, whereas the Sabine absorption coefficient $\overline{\alpha}$ is a measured quantity (see Table 1).

The preceding considerations of the reverberation time are based upon the assumption that the reverberant field is diffuse. However, enclosures are often encountered in which the field is obviously not diffuse. In this case a proposal due to Fitzroy has had some success.[9] Fitzroy proposes a modification to the Norris Eyring equation (20) as

$$T_{60} = \frac{55.25V}{-Sc} \sum_{i=1}^{3} \frac{S_i}{S \log_e(1 - \alpha_i)}. \qquad (25)$$

In Eq. (25), S_i and α_i $(i = 1-3)$ are respectively the total side-wall area S_{sw} and mean side-wall absorption coefficient α_{sw}, the total end-wall area S_{ew} and mean end-wall absorption coefficient α_{ew}, and the total floor and ceiling area S_{fc} and mean floor and ceiling absorption coefficient α_{fc}. The total room area $S = S_{\text{sw}} + S_{\text{ew}} + S_{\text{fc}}$. The equation should be used with caution as it has only been shown to provide reliable results for the cases considered by Fitzroy and there are other cases for which it is incorrect.

Optimum values of the reverberation times for various uses of an enclosed space may be calculated as[10]

$$T_{60} = K(0.0118V^{1/3} + 0.1070). \qquad (26)$$

The constant K of Eq. (26) takes the following values according to the proposed use: For speech $K = 4$, for orchestra $K = 5$, and for choirs and rock bands $K = 6$. At frequencies in the 250-Hz octave band and lower frequencies an increase over the value calculated by Eq. (26), ranging from 40% at 250 Hz to 100% at 63 Hz, is suggested.

5 MEASUREMENT OF THE SABINE ABSORPTION COEFFICIENT

Equation (15) forms the basis for the definition of the Sabine absorption coefficient $\overline{\alpha}$ and its measurement. It is important to note that it is the result of several approximations, and consequently it should come as no surprise that values of $\overline{\alpha}$ are sometimes determined that are greater than unity, suggesting that the material tested is capable of absorbing more sound energy than is incident

TABLE 1 Octave Band Average Sabine Sound Absorption Coefficients for Some Commonly Used Materials

Material	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
100% occupied audience (orchestra and chorus areas); upholstered seats	0.34	0.52	0.68	0.85	0.97	0.93	0.85
Unoccupied; average well-upholstered seating areas	0.28	0.44	0.60	0.77	0.89	0.82	0.70
Unoccupied; leather-covered upholstered seating areas		0.40	0.50	0.58	0.61	0.58	0.50
Wooden pews; 100% occupied		0.57	0.61	0.75	0.86	0.91	0.86
Wooden chairs; 100% occupied		0.60	0.74	0.88	0.96	0.93	0.85
Wooden chairs; 75% occupied		0.46	0.56	0.65	0.75	0.72	0.65
Fiberglass or rockwool blanket 60 kg/m^3							
25 mm thick (typical)		0.18	0.24	0.68	0.85	1.00	1.00
50 mm thick (typical)		0.25	0.83	1.00	1.00	1.00	1.00
(consult manufacturer)							
Floors							
Wood platform with large space beneath		0.40	0.30	0.20	0.17	0.15	0.10
Concrete or terrazzo		0.01	0.01	0.01	0.02	0.02	0.02
Linoleum, asphalt, rubber, or cork tile on concrete		0.02	0.03	0.03	0.03	0.03	0.02
Varnished wood joist floor		0.15	0.12	0.10	0.07	0.06	0.07
Carpet, heavy, on concrete		0.02	0.06	0.14	0.37	0.60	0.65
Carpet, heavy, on 40-oz (1.35 kg/m^2) hair felt or foam rubber		0.08	0.24	0.57	0.69	0.71	0.73
Carpet, 5 mm thick, on hard floor		0.02	0.03	0.05	0.10	0.30	0.50
Cork floor tiles ($\frac{3}{4}$ in. thick); glued down		0.08	0.02	0.08	0.19	0.21	0.22
Glazed tile/marble		0.01	0.01	0.01	0.01	0.02	0.02

upon it. When values of $\overline{\alpha}$ greater than unity are encountered, it is customary to arbitrarily set the measured quantity to unity.

Measured values of Sabine absorption coefficients for some commonly used acoustic materials are given in Table 1. More extensive tables are available in the literature[11–13].

In presenting absorption coefficient information, the term *noise reduction coefficient* (NRC) is sometimes used.[11] The latter quantity is defined as the arithmetic average of absorption coefficients determined at 250, 500, 1000, and 2000 Hz. In such presentation it is customary to round off the average to the nearest multiple of 0.05.

The Sabine absorption coefficient of a material is measured using a reverberation chamber, in which reverberation decay times are determined, generally in one-third-octave bands, with and without the material under test. The differences in measured decay times with the absorbent material in place, T'_{60}, and absent, T_{60}, allow determination of the absorption due to the presence of the test material using the following equation [derived from Eq. (15)]:

$$\overline{\alpha} = \frac{55.3V}{Sc}\left(\frac{1}{T'_{60}} - \frac{S' - S}{S'T_{60}}\right). \tag{27}$$

In Eq. (27), S' is the total area of all room surfaces, including the sample when in place, and S is the area of material (usually between 10 and 12 m^2, with a width-to-length ratio of between 0.7 and 1.0) exposed to the sound field. As described in the relevant standards,[14–17] the edges of the sample must be isolated from the sound field using appropriate rigid edging. An absorption coefficient is calculated using Eq. (27) for each one-third-octave band.

It is generally recognized that the Sabine equation

TABLE 1 (*Continued*)

Material	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Walls							
Plush curtain, deeply folded		0.15	0.45	0.90	0.92	0.95	0.95
Polyurethane foam, 27 kg/m^3, 15 mm thick		0.08	0.22	0.55	0.70	0.85	0.75
Acoustic plaster, 10 mm thick, sprayed on solid wall		0.08	0.15	0.30	0.50	0.60	0.70
Hard surfaces (brick walls, plaster, hard floats, etc.)		0.02	0.02	0.03	0.03	0.04	0.05
Slightly vibrating walls (suspended ceilings, etc.)		0.10	0.07	0.05	0.04	0.04	0.05
Strongly vibrating surface (wooden paneling over airspace, etc.)		0.40	0.20	0.12	0.07	0.05	0.05
Plaster, gypsum or lime, smooth finish							
On brick		0.013	0.015	0.02	0.03	0.04	0.05
On concrete block		0.012	0.09	0.07	0.05	0.05	0.04
On lath		0.014	0.10	0.06	0.04	0.04	0.03
Curtains							
Light velour, 10 oz/yd^2, hung	0.03	0.04	0.11	0.17	0.24	0.35	
Medium velour, 14 oz/yd^2 draped to half area	0.07	0.31	0.49	0.75	0.70	0.60	
Heavy velour, 18 oz/yd^2 draped to half area	0.14	0.35	0.55	0.72	0.70	0.65	
Glass							
Glass, heavy plate	0.18	0.06	0.04	0.03	0.02	0.02	
Ordinary window	0.35	0.25	0.018	0.12	0.07	0.04	
Stage openings	0.30	0.40	0.50	0.60	0.60	0.50	
Water (surface of pool)	0.01	0.01	0.01	0.015	0.02	0.03	

(15) is a gross simplification of a complex problem. However, it has recently been proposed[5] that, in fact, the Sabine equation is exact and that, when edge diffraction is taken into account in the computation of Eq. (21), the Sabine absorption coefficient of Eq. (15) is equal to the statistical absorption coefficient of Eq. (21). The effect of edge diffraction is to increase the effective area of the material.[1] Unfortunately, edge diffraction is ignored both in the standard measurement of Sabine absorption coefficients and in their application. Currently available procedures for the computation of the Sabine absorption coefficient are not adequate for general use.[18]

One problem resulting from neglecting edge diffraction in the standard test is that absorption coefficients greater than 1 are often obtained, whereas in principle only values less than 1 are permitted by the basic assumption that the absorption coefficient is a measure of the incident energy that is absorbed. A second problem is that results for the same test material can be highly variable. The basic problem remains that no theoretical solution has been accepted relating statistical absorption to reverberation decay in a Sabine room, although one has recently been proposed.[5] However, conditions necessary for agreement among results for the determination of the Sabine absorption coefficient of the same test material tested in several rooms have been determined.[19] Reverberation chambers ranging in size from 106 to 607 m^3 have been shown to give essentially the same absorption coefficients for a 10.5-m^2 test sample of Silan when measured using the reverberation decay method, and the results are interpreted using Sabine's equation provided that the reverberation chamber contains auxiliary diffusing elements of total area (both sides) greater than 1.4 times the chamber floor area. The problem of variability of test results can be overcome with adequate auxiliary diffusion.

In the same series of investigations test samples ranging in size from 22.5 to 10.5 m^2 in the large 607-m^3 room and from 10.5 to 5.0 m^2 in the small 106-m^3 room all gave similar results. Absorption coefficients greater than 1 in the frequency range of the 500-Hz one-third-octave band and higher frequency bands were reported, although the predicted statistical absorption coefficient has a maximum value of 0.954. Only a small tendency for the smallest test sample to be associated with the greatest sound absorption coefficient is to be observed in the range of 500 Hz in the largest room test data and in data of an intermediate-size 208-m^3 room. The tendency is not observed in the data of the smallest room. It is to be noted that an edge effect would predict that greater absorption coefficients would be observed in smaller test samples tested in the smallest room, contrary to what has been reported.

In rooms with inadequate auxiliary diffusion a decrease in effective absorption is observed at high frequencies and a tendency for smaller test samples to give larger absorption coefficients than larger samples is quite evident. However, the latter effects tend to disappear or to be greatly reduced when there is adequate auxiliary diffusion in the room. Alternatively, diffraction at the test sample edges, which is ignored in the standard test and has been assumed to account for the greater effectiveness of small test samples compared to large samples, seems to be overridden when diffusion is adequate. Adequate auxiliary diffusion seems to suppress any edge effects attributed to diffraction.

If the reverberation room is considered to be a model for all Sabine-type rooms, then the problems of data interpretation, which have been described, are equally true for application of measured absorption coefficient data to the design of Sabine-type rooms. Until reliable analysis is provided, Sabine absorption coefficient data must be used with care.

6 EFFECT OF AN AUDIENCE

In occupied auditoria the audience will represent the major contribution to sound absorption. In considering the effect of an audience upon the sabins of absorption they contribute, two possibilities have been investigated. One method is for some effective area of total absorption to be attributed to a person, for example, some number of sabins (square metres) per person and the area per person multiplied by the expected total number of people, to obtain an estimate of the expected absorption. Alternatively, an effective absorption coefficient may be attributed to the area occupied by the expected audience. The latter procedure has been shown to give more accurate results than the former, which can be in serious error. Beranek has investigated this problem and finds an effective absorption coefficient of about 0.92 for the 500- and 1000-Hz octave bands, giving an effective absorption in sabins of between 2.3 and 5.7 m^2 per occupied seat.[11]

7 AIR ABSORPTION

In addition to energy loss on reflection at the boundaries of an enclosure, some energy is dissipated while the sound wave propagates through the air. Such propagational loss is small and is generally only important in very large rooms and at frequencies above about 500 Hz. In considering the absorption of sound as it travels through air, it will be assumed that the fraction of sound energy lost during propagation is linearly related to the mean free path between reflections. As may readily be shown,[3] the total mean Sabine absorption coefficient becomes

$$\overline{\alpha} = \overline{\alpha}_w + 9.21 \times 10^{-4} mV/S. \tag{28}$$

Values of the absorption of sound m expressed in decibles per thousand meters as a function of frequency and relative humidity, given in Table 2, were calculated using Sutherland's method.[20] In Eq. (28) the Sabine absorption coefficient for the boundary of a room has been identified with a subscript w. It is to be noted that what is implied by the mean Sabine absorption coefficient $\overline{\alpha}$ in Eq. (15) is the quantity given by Eq. (28).

8 STATISTICAL SOUND ABSORPTION COEFFICIENT

In a diffuse sound field, sound energy will be incident upon a boundary from all directions in the adjacent half space. As the sound absorbed at the boundary can be expected to be dependent upon the angle of incidence, it is convenient to define a statistical sound absorption coefficient α_{st}, which is the average value of sound absorption coefficient, assuming all angles of incidence with equal intensity to be equally probable. The statistical sound absorption coefficient, together with Eq. (20), is recommended for use in the design of large auditoria in place of the Sabine absorption coefficient and Eq. (15).

In principle, the Sabine absorption coefficient should be the same as the statistical sound absorption coefficient, but this is not the case. Though similar in concept, the former is defined by Eq. (15) and the associated measurement of Eq. (27), which is not necessarily the definition of the latter. In fact, no way of making

TABLE 2 Attenuation due to Atmospheric Absorption[a]

Relative Humidity (%)	Temperature (°C)	m (dB/1000 m)							
		63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
25	15	0.2	0.6	1.3	2.4	5.9	19.3	66.9	198.0
	20	0.2	0.6	1.5	2.6	5.4	15.5	53.7	180.5
	25	0.2	0.6	1.6	3.1	5.6	13.5	43.6	153.4
	30	0.1	0.5	1.7	3.7	6.5	13.0	37.0	128.2
50	15	0.1	0.4	1.2	2.4	4.3	10.3	33.2	118.4
	20	0.1	0.4	1.2	2.8	5.0	10.0	28.1	97.4
	25	0.1	0.3	1.2	3.2	6.2	10.8	25.6	82.2
	30	0.1	0.3	1.1	3.4	7.4	12.8	25.4	72.4
75	15	0.1	0.3	1.0	2.4	4.5	8.7	23.7	81.6
	20	0.1	0.3	0.9	2.7	5.5	9.6	22.0	69.1
	25	0.1	0.2	0.9	2.8	6.5	11.5	22.4	61.5
	30	0.1	0.2	0.8	2.7	7.4	14.2	24.0	68.4

[a] Calculated using Sutherland's method (Ref. 20).

a direct measurement of the statistical sound absorption coefficient of a material has been demonstrated, though predicted dependence upon angle of incidence has been experimentally verified. On a theoretical basis, it may be shown that the statistical sound absorption coefficient α_{st} has a maximum value of 0.954.

The statistical sound absorption coefficient of porous materials may be calculated based upon either a measurement of the normal impedance using an impedance tube or a measurement of the material flow resistance.[3,21] Both methods will be reviewed and the results of some calculations will be presented in the following sections.

8.1 Impedance Tube Method

Any heavy-walled tube may be used to construct an impedance tube. A source of sound is placed at one end and the sample to be investigated at the other. Investigation is limited to the frequency range in which plane-wave propagation in the tube may be assured. For a circular cross-sectional tube of diameter d, this frequency range is between zero and an upper frequency defined by $f = 0.586c/d$. The normal impedance of the sample may be determined from investigation of the sound field in the tube. Test methods are described in various standards.[22,23]

When the material under test is in place and the sound field in the tube is explored, a series of maxima and minima will be observed. The maxima $L_{\max}$ will be effectively constant in level, but the minima will be observed to increase in level as the mesurement point is moved away from the sample material. Straight-line extrapolation of the minima back to the surface of the sample is recommended to determine an estimate of the minimum level $L_{\min}$.[1] The standing-wave ratio L_0 is given as

$$L_0 = L_{\max} - L_{\min}. \tag{29}$$

The location D of the first minimum relative to the surface of the sample is determined. The minima are located one half-wavelength ($\lambda/2$) apart, from which the ratio D/λ may be formed. The statistical sound absorption coefficient may then be calculated as

$$\alpha_{\mathrm{st}} = \frac{8\cos\beta}{\xi} \times \left[1 - \frac{\cos\beta}{\xi}\log_e(1 + 2\xi\cos\beta + \xi^2) + \frac{\cos 2\beta}{\xi\sin\beta}\tan^{-1}\frac{\xi\sin\beta}{1+\xi\cos\beta}\right], \tag{30}$$

where

$$K_0 = 10^{L_0/20}, \tag{31a}$$

$$M = 0.5(K_0 + K_0^{-1}), \tag{31b}$$

$$N = 0.5(K_0 - K_0^{-1}), \tag{31c}$$

$$\theta = 360\left(\frac{2D}{\lambda} - 0.5\right)\ \mathrm{deg} \tag{32a}$$

$$R = (M - N\cos\theta)^{-1}, \tag{32b}$$

$$X = RN\sin\theta; \tag{32c}$$

$$\xi = (R^2 + X^2)^{1/2}, \tag{33a}$$

$$\beta = \tan^{-1}\frac{X}{R} \tag{33b}$$

8.2 Flow Resistance Measurement Method

If a constant differential pressure is imposed across a layer of bulk porous material of open cell structure, a steady flow of gas will be induced through the material. Provided that the flow velocity is small, the differential pressure P and the induced normal velocity U of gas at the surface of the material (volume velocity per unit surface area) are linearly related for a wide range of materials. The ratio of differential pressure in pascals to normal velocity in metres per second is known as the flow resistance R_f (MKS rayls) of the material.

If the material is generally of uniform composition, then the flow resistance is proportional to the material thickness L, and the flow resistivity may be defined as $R_1 = R_f/L$. The measurement of flow resistance is described in a standard[21] and may be calculated for a sample of area A from the measured quantities as

$$R_f = \frac{PA}{U}. \tag{34}$$

As has been shown, the quantities R and X of Eqs. (32b) and (32c) may be calculated in terms of the dimensionless parameter $\rho f/R_1$, where ρ is the density of the gas (air) and the other quantities have been defined. By this means, Fig. 2 has been prepared. It is found that the calculated statistical sound absorption coefficient is close to the optimal value of 0.95 when the thickness of material is greater than 0.1 wavelengths and the material flow resistance is between $2\rho c$ and $5\rho c$.

The effective thickness of the absorptive material may be increased by providing a backing cavity equal in depth to the thickness of the porous material.[24] In this case, the composite of material and cavity behaves similar to a layer of material twice as thick. No advantage is gained

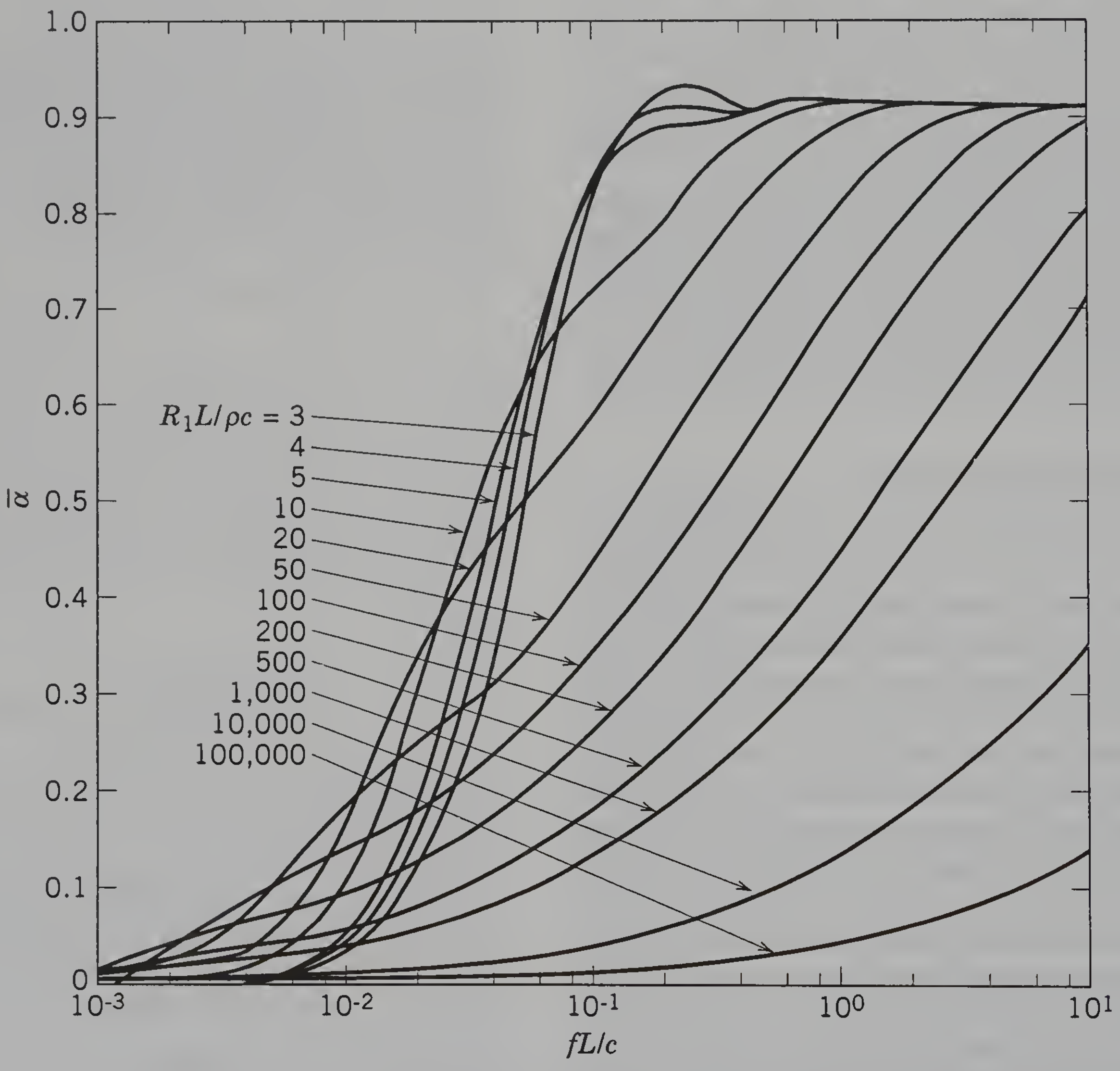


Fig. 2 Calculated statistical absorption coefficient for a porous surface in a reverberant field as a function of a frequency parameter for various indicated values of normalized flow resistance: R_1L, flow resistance of porous material (MKS rayls); ρ, density of air (kg/m^3), c, speed of sound in air (m/s); f, frequency (Hz) of incident sound; L, porous material thickness (m).

with a deeper backing cavity, and care must be taken in its construction to provide septa to prevent propagation in the cavity parallel to the material surface, which will degrade the performance of the treatment.

9 EFFECT OF PROTECTIVE FACINGS ON ABSORBING MATERIAL

When mechanical protection is needed for the porous liner, it may be covered using perforated wood, plastic, or metal panels. If the open area provided by the perforations is greater than about 25%, the expected absorption is entirely controlled by the properties of the porous liner, and the panel has no effect.

Alternatively, if the facing panel has an open area of less than 25%, then the frequency at which absorption is maximum may be calculated by treating the facing as an array of Helmholtz resonators. This procedure leads to the approximate equation

$$f_{\max} = \frac{c}{2\pi}\left[\frac{P/100}{L[t + 0.85d(1 - 0.22d/q)]}\right]^{1/2}, \qquad (35)$$

where P is the percentage open area of the panel, L is the depth of the backing cavity including the porous material layer, t is the panel thickness, d is the diameter of the perforations, and q is the spacing between holes. If the porous material fills the entire cavity so that the thickness of the porous material is also L, then the speed of sound c should be replaced with $0.85c$ to account for isothermal rather than adiabatic propagation of sound in the porous material at low frequencies. The condition $fL/c < 0.1$ must be satisfied for this equation to give results with less than 15% error. Measured data for a panel of 10% open area are presented in Fig. 3.

If a perforated facing is used over a porous liner wrapped in an impervious membrane, care must be taken to ensure that the facing and liner are not in contact; otherwise, the sound-absorbing properties of the combination will be severely degraded. Noncontact is usually achieved in practice by placing a 12-mm-square wire mesh spacer between the liner and perforated sheet.

10 MODALLY REACTIVE BOUNDARIES

Modal coupling has received attention in the literature concerned with sound level prediction in vehicles but generally not in rooms. However, one case of some importance for the consideration of sound absorption in rooms has been considered and may be reported here.

10.1 Panel Absorption: Analytical Approach

When low-frequency absorption is required, panel absorbers are often the best option. To achieve absorp-

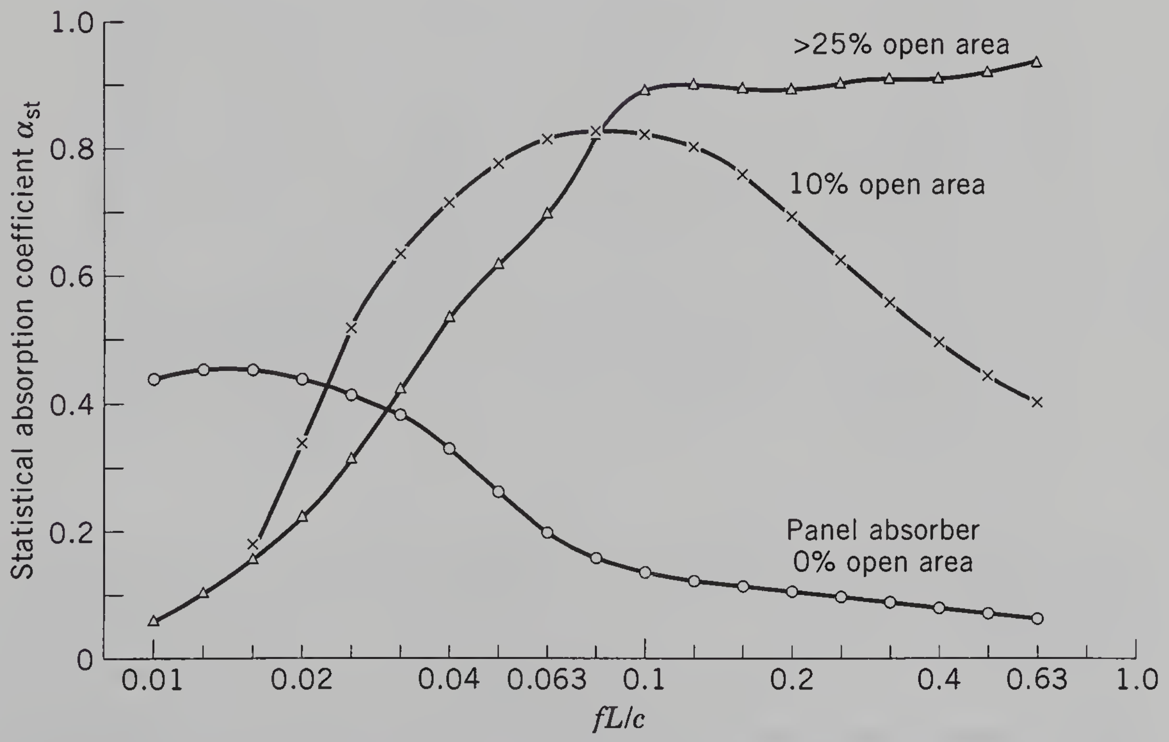


Fig. 3 Effect of perforations on the sound absorption of a plywood panel, 3 mm thick and 2.5 kg/m^2 surface weight, backed by a cavity 50 mm thick and filled with a flexible porous liner of flow resistance $5\rho c$.

tion, panels are mounted spaced out from the wall providing a shallow backing cavity and resonant response in the frequency range of interest. Sometimes the cavity is filled with a porous material to improve the absorption.

The Sabine absorption coefficient of n resonant panels of total surface area S and individual surface area A_p may be calculated by using the following formulas.[25] It is to be noted that in this case the Sabine absorption coefficient is explicitly a function of the properties of the room as well as the properties of the panel, and consequently fairly good results can be expected in using the formula. Unfortunately, the price paid for good results will be quite a few measurements to determine the properties of both room and panels:

$$\bar{\alpha} = \frac{4V\pi f}{Sc}\left\{\left|\left(\eta_A + \frac{n_p}{n_A}\,\eta_{pA} + \eta_{pT}\right) - \left[\left(\eta_A + \frac{n_p}{n_A} - \eta_{pT}\right)^2 + 4\,\frac{n_p}{n_A}\,\eta_{pA}^2\right]^{1/2}\right| - 2\eta_A\right\}, \tag{36}$$

where the quantities in the above equation are as follows. The room loss factor with the panel absent is

$$\eta_A = \frac{2.20}{fT_{60A}}. \tag{37}$$

The panel loss factor as measured in an anechoic room is

$$\eta_{pT} = \frac{2.20}{fT_{60p}}. \tag{38}$$

The panel coupling loss factor is

$$\eta_{pA} = \frac{\rho c \sigma_{\mathrm{rad}}}{2\pi m_p}. \tag{39}$$

The panel modal density is

$$n_p = \frac{\sqrt{3}A_p}{c_L h}. \tag{40}$$

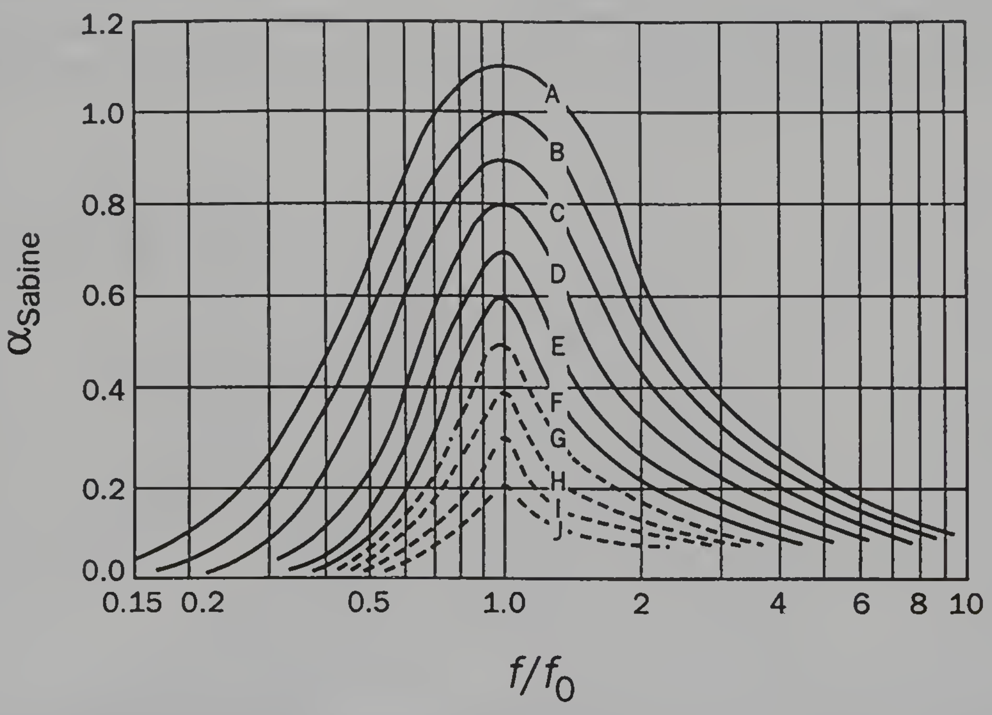


Fig. 4 Sabine absorption coefficients for resonant plywood panels. The panel configurations corresponding to the curves labeled *A–J* may be identified using Fig. 5. Dashed curves (*G–J*) represent configurations with no absorptive material in the cavity behind the panel. Configurations *A–F* require a sound-absorbing blanket between the panel and backing wall. The blanket must not contact the panel and should be between 10 and 50 mm thick and consist of glass or mineral fiber with a flow resistance between 1000 and 2000 MKS rayls. Panel supports should be at least 0.4 m apart. Courtesy of Hardwood Plywood & Veneer Association.

The surface area of all panels is

$$S = \sum_{i=1}^{n} A_{pi}. \tag{41}$$

The modal density n_A for any room may be calculated using Eq. (1). The quantities T_{60A} and T_{60p} are, respectively, the reverberation times of the room (without panels) and the panels (in a semianechoic space), c_L is the longitudinal wave speed in the panel, f is the center frequency of the frequency band of interest, h is the panel thickness, ρ is the density of air, and m_p is the mass per unit surface area of the panel.

The panel radiation ratio σ_{rad} at frequency f used in the preceding equations may be calculated by using the equations

$$\sigma_{\text{rad}} = 0.367\gamma \frac{c_L}{c}\frac{Ph}{A_p} \sin^{-1}\sqrt{\frac{f}{f_c}}, \qquad f < f_c, \tag{42}$$

$$\sigma_{\text{rad}} \simeq \sqrt{\frac{Pf_c}{2c}}, \qquad f \simeq f_c, \tag{43}$$

$$\sigma_{\text{rad}} = \sqrt{1 - \frac{f_c}{f}}, \qquad f > f_c, \tag{44}$$

$$f_c = 0.551\frac{c^2}{c_L h}. \tag{45}$$

In the preceding equations the quantities are the panel perimeter P and area A_p. The panel is assumed to be isotropic of uniform thickness h and longitudinal wave speed c_L. For steel and aluminum c_L takes the value of about 5200 m/s, while for wood the value lies between 3800 and 4500 m/s. For simply supported panels γ takes the value of 1 while for clamped edge panels γ takes the value of 2. All other conditions will lie between these extremes.

10.2 Panel Absorption: Empirical Approach

A less accurate but much simpler empirical prediction scheme for determining the Sabine absorption coefficient is presented in Figs. 4 and 5. First the Sabine absorption required is selected from curves *A–J* in Fig. 4. The solid curves refer to panels for which the backing cavity contains a porous material 25 mm thick, with a flow resis-

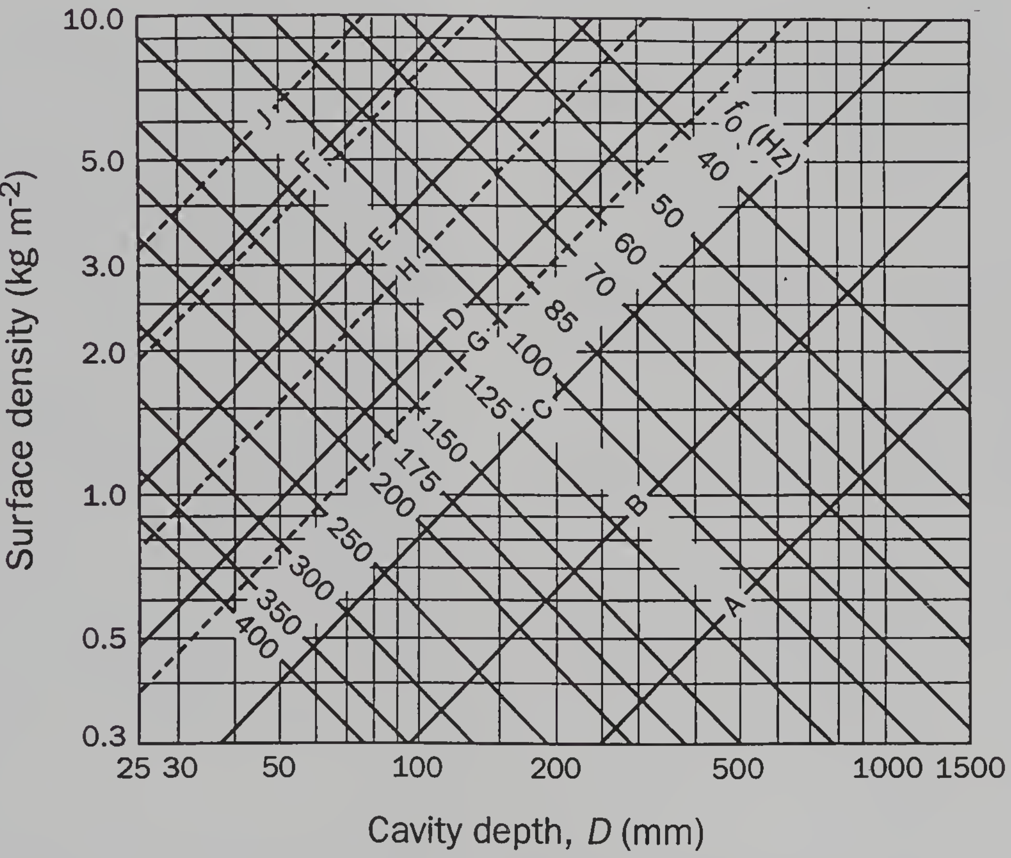


Fig. 5 Design curves for resonant plywood panels, to be used in conjunction with Fig. 4. The quantity f_0 is the frequency at which maximum sound absorption is required. Courtesy of Hardwood Plywood & Veneer Association.

tance of between $2\rho c$ and $5\rho c$, while the dashed curves refer to panels for which there is no porous material in the backing cavity. The frequency at which maximum absorption is required, f_0, is then used to enter Fig. 5, from which the required depth of backing cavity and surface density of the panel may be determined.

Acknowledgments

The authors thank Fergus Fricke for helpful discussions and contructive criticisms.

REFERENCES

1. P. M. Morse and R. H. Bolt, "Sound Waves in Rooms," *Rev. Modern Phys.*, Vol. 16, 1944, pp. 65–150.
2. L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, Springer-Verlag, New York, 1973.
3. D. A. Bies and C. H. Hansen, *Engineering Noise Control: Theory and Practice*, 2nd. ed., E. and F. N. Spon, London, 1996.
4. W. H. Sabine, *Collected Papers on Acoustics*, Dover, New York, 1964.
5. D. A. Bies, "Notes on Sabine Rooms," *Acoust. Austral.*, Vol. 23, 1995, pp. 97–103.
6. C. F. Eyring, "Reverberation Time in 'Dead Rooms,' " *J. Acoust. Soc. Am.*, Vol. 1, 1930, pp. 217–41.
7. G. Millington, "A Modified Formula for Reverberation," *J. Acoust. Soc. Am.*, Vol. 4, 1932, pp. 69–82.
8. W. J. Sette, "A New Reverberation Time Formula," *J. Acoust. Soc. Am.*, Vol. 4, 1933, pp. 193–210.
9. D. Fitzroy, "Reverberation Formula Which Seems to be More Accurate with Nonuniform Distribution of Absorption," *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 893–97.
10. R. W. B. Stephens and A. E. Bate, *Wave Motion and Sound*, Edward Arnold, London, 1950.
11. M. J. Crocker and A. J. Price, *Noise and Noise Control*, CRC Press, Cleveland, OH, 119 Volume I, 1975, and Volume II, 1985.
12. H. Kuttruff, *Room Acoustics*, 2nd ed., Applied Science, London, 1979.
13. L. J. Doelle, *Environmental Acoustics*, McGraw-Hill, New York, 1972.
14. International Standardization Organization, "Measurement of Sound Absorption in a Reverberation Room," ISO R354-1963, 1963.
15. Australian Standard, "Measurement of Sound Absorption in a Reverberation Room," AS 1045-1988, 1988.
16. American Society for Testing and Materials, "Test Method for Sound Absorptoin and Sound Absorption Coefficients by the Reverberation Room Method," ASTM C423-81a, 1981.
17. American Society for Testing and Materials, "Practice for Mounting Test Specimens during Sound Absorption Tests," ASTM E795-81, 1981.
18. T. D. Northwood, "Absorption of Diffuse Sound by a Strip or Rectangular Patch of Absorptive Material," *J. Acoust. Soc. Am.*, Vol. 38, 1963, pp. 1173–1177.
19. W. A. Davern and P. Dubout, "First Report on Australian Comparison Measurements of Sound Absorption Coefficients," CSIRO—Division of Building Research, Highett, Victoria, Australia, 1980.
20. J. Piercy, T. Embleton, and L. Sutherland, "Review of Noise Propagation in the Atmosphere," *J. Acoust. Soc. Am.*, Vol. 61, 1977, p. 1403.
21. American Society for Testing and Materials, "Test Method for Airflow Resistance of Acoustical Materials," ASTM C522-80, 1980.
22. American Society for Testing and Materials, "Test Method for Impedance and Absorption of Acoustic Materials by the Impedance Tube Method," ASTM C384-77, 1977.
23. Australian Standard, "Method for Measurement of Normal Incidence Sound Absorption Coefficient and Specific Normal Acoustic Impedance of Acoustic Materials by the Tube Method," AS 1935–1976, 1976.
24. D. A. Bies and C. H. Hansen, "Flow Resistance Information for Acoustical Design," *Appl. Acoust.*, Vol. 13, 1980, pp. 357–391.
25. J. Pan and D. A. Bies, "The Effect of Fluid-Structural Coupling on Acoustical Decays in a Reverberation Room in the High-Frequency Range," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 718–27.

93

SOUND INSULATION: AIRBORNE AND IMPACT

A. C. C. Warnock and Wolfgang Fasold

1 INTRODUCTION

Problems of sound propagation in buildings can be considered as the classic questions of architectural acoustics. Although early publications date back to the beginning of the twentieth century,[1] theoretical solutions sufficient to describe the sound insulation of building elements and of buildings did not appear until between 1930 and 1950.[2–6]

In buildings, it is customary, to differentiate between airborne and structure-borne sound excitation. Airborne sound originates from a source radiating sound waves into the air that then impinge on the surfaces of building elements. Examples are human speech or music from a loudspeaker. Problems of airborne sound insulation occur between adjacent rooms, between neighboring buildings, and with respect to outdoor sound sources entering the building.

Structure-borne sound is generated by direct mechanical excitation of building elements, for example, by machines and by people walking. Footstep noise is an example of impact sound, a subset of structure-borne sound. This chapter only deals with impact sound, not with structure-borne sound in general. (See Chapters 64 and 98.) Problems of impact sound insulation primarily occur with floors and stairs in apartment houses, schools, hospitals, and hotels. While perhaps not so frequent, problems can also arise with slamming doors in cupboards and in hallways.

2 AIRBORNE SOUND INSULATION

2.1 Sound Reduction Index or Sound Transmission Loss

Sound transmission through a wall separating two adjacent rooms is shown in principle in Fig. 1.[7] The sound power W_1 that impinges on the surface of the wall in room 1 (the source room) partly reflects back into the room, partly propagates to other connecting structures (W_3 and W_4), partly dissipates as heat within the material of the wall, and partly transmits into room 2 (the receiving room) with sound power W_2.

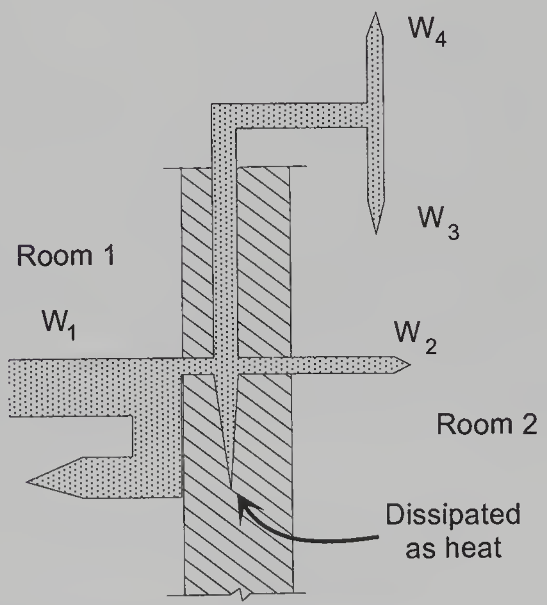


Fig. 1 Principle of sound transmission through a wall: W_3 and W_4 represent flanking sound transmitted to other parts of the structure; W_3 is eventually radiated into room 2, W_4 is not.

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

The sound transmission coefficient τ is defined as

$$\tau = \frac{W_2}{W_1}. \qquad (1)$$

Typically, sound transmission coefficients are very small numbers, and it is more convenient to use the sound reduction index R, which is defined as

$$R = 10 \log \frac{1}{\tau} \quad \text{dB}. \qquad (2)$$

The sound reduction index is also known as sound transmission loss (TL), and both terms are used in this chapter. Sound transmission varies with the frequency of the incident sound waves, and measurements are usually made in one-third-octave bands.

2.2 Airborne Sound Reduction Measurement

Methods for measuring sound insulation are specified in national and international standards. The most common method of determining airborne sound insulation is the two-room method.[8–10] These measurements are performed in laboratory test facilities in which the only path for the transmission of sound is through the specimen; transmission along other paths is suppressed.[11] In some cases, however, testing is carried out with well-defined flanking conditions[12] that are appropriate for the building types used in a region. *Flanking transmission* is transmission along some path other than that directly through the main partition. It is discussed further in Section 2.15. In Fig. 1 W_3 is an example of flanking transmission.

The test specimen is mounted between a source room and a receiving room, both of which are hard surfaced and reverberant. In the source room an approximately diffuse sound field is generated using loudspeakers, usually fed with white noise. The average sound pressure level is measured in each room in one-third-octave bands. The sound pressure level in the receiving room is determined by the sound reduction index (transmission loss) R of the test specimen, the area of the test specimen, and the equivalent sound absorption area of the receiving room. The relationship between the measured quantities is

$$R = L_1 - L_2 + 10 \log \frac{S}{A} \quad \text{dB}, \qquad (3)$$

where L_1 is the sound pressure level in the source room and L_2 the sound pressure level in the receiving room (both in decibels) and S is the area of the test specimen and A the equivalent sound absorption area (also called sound absorption) of the receiving room (in square metres).

Here, A can be determined from

$$A = \frac{0.921 V d}{c} \quad \text{m}^2, \qquad (4)$$

or from

$$A = 0.16 \frac{V}{T} \quad \text{m}^2, \qquad (5)$$

where V is the volume of the receiving room (in cubic metres), d is the rate of decay of the sound pressure level in the room (in decibels per second), $T = 60/d$ is the reverberation time of the receiving room (in seconds), and c is the speed of sound in air (in metres per second). The velocity of sound in air is temperature dependent and is given by $c = 20.047\sqrt{273.15 + t}$ metres per second, where t is temperature in degrees Celsius.

Laboratory methods are used to measure the sound insulation of building elements for comparison with building code requirements and to allow the comparison of different systems. Similar methods are also available to check the quality of sound insulation achieved in buildings.[13,14] In the International Standardization Organization (ISO) building acoustics standards a prime after a symbol, for example R'_w, means that flanking sound transmission is either suspected or known to be present.

2.3 Sound Intensity Measurements

Since 1981,[15] sound intensity measurements have been used to determine sound insulation. The use has been standardized[16] but not widely. Many researchers[17,18] have found differences at low frequencies between results from the standard two-room test procedures and those from intensity measurements. The test procedure consists of measuring the average sound pressure in a room on one side of a test specimen, $\langle p^2 \rangle$, and then measuring the average sound intensity over the surface of the specimen on the other side, $\langle I \rangle$. The method offers the possibility of obtaining the sound reduction index without the influence of flanking transmission in laboratories or in buildings, but great care is needed to obtain accurate results.

In a diffuse sound field, the intensity incident on a specimen is $\langle p^2 \rangle / 4\rho c$. Thus the transmission coefficient is given by

$$\frac{1}{\tau} = \frac{\langle p^2 \rangle}{4\rho_0 c \langle I \rangle}, \qquad (6)$$

or in decibels,

$$R = L_p - L_I - 6 \text{ dB}, \tag{7}$$

where L_p is the sound pressure level re 2×10^{-5} Pa and L_I the sound intensity level re 10^{-12} W/m^2 (both in decibels).

2.4 Short Test Methods

Short test methods for airborne and impact sound insulation of building elements have lost their importance since advanced instrumentation is available to simplify measuring and evaluating procedures. Methods that have been used in practice are based on the fact that an approximate single number rating can be obtained by measuring a frequency-weighted sound pressure level in the receiving room. The weighting curve depends on the spectrum of the sound source. Problems arise because formulas defining the sound reduction index and normalized impact sound pressure level (see Sections 2.1 and 3.1) require knowledge of the reverberation time in the receiving room. Therefore, such short test methods may be efficient only in cases in which an estimate of reverberation time is permissible.

2.5 Facade Sound Reduction

To measure the airborne sound insulation of facade elements and facades,[19] special methods are in use. For measurement with loudspeaker noise, the loudspeaker is placed outside the building at a distance of about 10 m radiating at an angle of about 45° to the test specimen. Another measuring method especially suitable for buildings near roads with high traffic density uses the traffic noise that impinges on the surface of the test specimen. If the test specimen can be fully opened (e.g., window or door) and is mounted in a heavy wall, the airborne sound insulation can be determined by measuring the average sound pressure level in the receiving room with the specimen in the open and in the closed position.

2.6 Single-Number Ratings

Examples of sound transmission measurements for some common building materials are shown in Fig. 2. Such plots contain more information than is convenient for use in building codes. Methods have been standardized whereby the frequency-dependent values of airborne sound insulation can be converted into a single-number rating by means of a reference curve.[20–22] These rating systems can be applied to many different types of sound reduction measurements, not just transmission loss or sound reduction index.

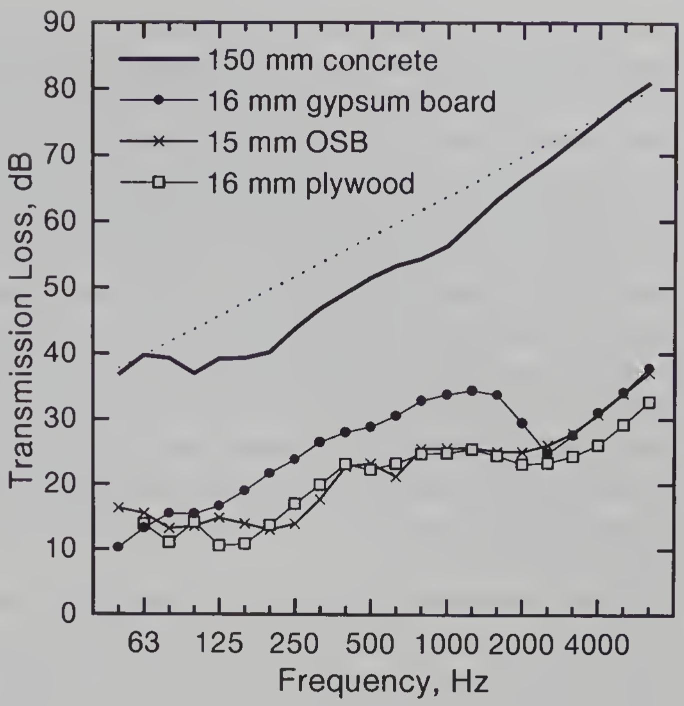


Fig. 2 Examples of sound transmission through some common building materials: 150-mm concrete (345 kg/m^2), 16-mm plywood (7.5 kg/m^2), 16-mm gypsum board (11.3 kg/m^2), 15-mm oriented strand board (OSB, 8.8 kg/m^2). Dashed line represents mass law predictions for 345 kg/m^2.

The ISO and American Society for Testing and Materials (ASTM) reference curves are shown in Fig. 3 fitted to measured curves. To evaluate the results of a measurement, the ISO reference curve (100–3150 Hz) is shifted

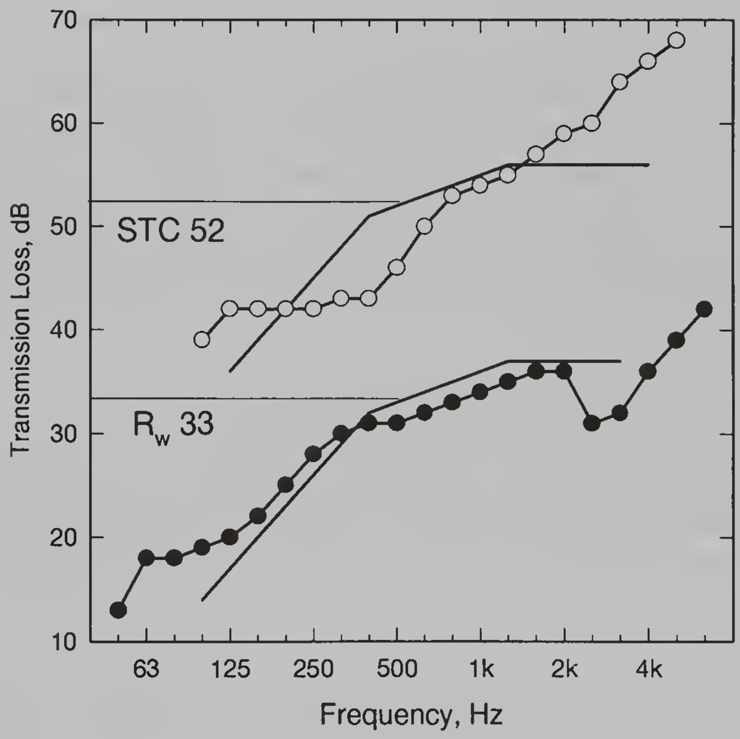


Fig. 3 ISO and ASTM curves of reference values for airborne sound compared with measurement results: (●) two layers of 13-mm gypsum board (20 kg/m^2); (○) 100-mm concrete (230 kg/m^2).

in steps of 1 dB toward the measured curve until the mean unfavorable deviation is as large as possible but not more than 2 dB. An unfavorable deviation at a particular frequency occurs when the measured result is less than the reference value. The mean unfavorable deviation is calculated by dividing the sum of the unfavorable deviations by 16, the total number of measurement frequencies. The value, in decibels, of the reference curve at 500 Hz after shifting it according to this procedure is denoted the *weighted sound reduction index* R_w. (In ISO terminology, a subscript W after a symbol means that the value presented is a weighted single-number rating of a one-third-octave band spectrum. The weighted sound reduction index R_w was formerly known as the airborne sound insulation index I_a.)

A similar procedure is used to calculate the *sound transmission class* (STC). Note that the frequency range of the STC reference curve is from 125 to 4000 Hz. During the evaluation procedure attention has to be paid to the maximum unfavorable deviation. If it exceeds 8 dB at any frequency, the ISO standard requires that it be recorded and reported. The ASTM standard does not allow any unfavorable deviation to exceed 8 dB. Once this value is attained, the reference curve can no longer be adjusted. Despite the difference in procedures, the numerical values of the STC and R_w are usually very close, although when the 8-dB rule controls the STC rating, differences of several decibels may occur. Where only one rating is given in this chapter, assume the value holds for both ratings.

Although the shapes of the reference curves were derived from a mean spectrum of noise inside dwellings, they have been also used for outdoor noise, especially for traffic noise; the standardized evaluation procedure is applied to exterior walls, windows, and doors as well as to interior partitions. This use provoked criticism and led to proposals for specific reference curves with a changed slope and frequency range that take into account the higher importance of low frequencies in the traffic noise spectrum.[23,24]

2.7 Single-Leaf Building Elements

The sound reduction index of a homogeneous, limp, nonporous plate for plane-wave incidence is given by the *mass law*[2,25] as

$$R = 10 \log \left[1 + \left(\frac{\pi f M \cos \theta}{\rho_0 c} \right)^2 \right] \quad \text{dB}, \qquad (8)$$

where f is frequency (in hertz); M is mass per unit area of the plate (in kilograms per square metre); θ is the angle of incidence; ρ_0 is the density of air, typically 1.2 kg/m^3; and c is the velocity of sound in air (in metres per second). For $t = 20°\text{C}$, $c = 343$ m/s and $\rho_0 c = 412\ \text{N} \cdot \text{s/m}^3$. The mass law predicts an increase in sound reduction index of about 6 dB for each doubling of mass per unit area or frequency.

Bending waves in panels with finite stiffness give rise to the *coincidence effect*.[5] The velocity of bending waves in panels increases with increasing frequency according to

$$c_B = \left(\frac{B\omega^2}{m} \right)^{1/4}, \qquad (9)$$

where B is the bending stiffness per unit width, $\omega = 2\pi f$ is the angular frequency (in reciprocal seconds), and m is the panel surface density (in kilograms per square metre). At every frequency above a certain *critical frequency* f_c there is an angle of incidence for which the wavelength of the bending wave becomes equal to the wavelength of sound in air projected on to the plate; the waves coincide. Because of the coincidence effect, a region of relatively low sound insulation occurs in some frequency range in the curve of the sound reduction index. For example, the curve for gypsum board in Fig. 2 shows a coincidence dip with a minimum at 2500 Hz. The critical frequency for a thin panel can be calculated from

$$f_c = \frac{c^2}{2\pi} \sqrt{\frac{\rho h}{B}} \quad \text{Hz}, \qquad (10)$$

where c is the velocity of sound in air (in metres per second), ρ is the density of material (in kilograms per cubic metre), h is the thickness of the panel (in metres), and B is the bending stiffness per unit width (in kilograms by square metres per square second). The bending stiffness is related to Young's modulus by

$$B = \frac{Eh^3}{12} (1 - \nu^2) \quad \text{kg} \cdot \text{m}^2/\text{s}^2, \qquad (11)$$

where E is Young's modulus (in newtons per square metre) and ν is Poisson's ratio. Thus for a given material the product $f_c h$ is a constant that depends only on the physical properties of the material. Table 1 gives values of this product for common building materials. The resonance dip due to the coincidence effect usually begins about an octave below the critical frequency. The depth of the resonance dip depends on the damping of the panel. Below the frequency range of coincidence the sound reduction index is determined by the mass law.

TABLE 1 Surface Mass for 1-mm Thicknesses and Constant $f_c h$ for Some Common Building Materials

Material	Surface Mass ρh(kg/m^2)	$f_c h$ (Hz · mm)
Aluminum	2.7	12,900
Concrete, dense poured	2.3	18,700
Hollow concrete block	1.1	20,900
Fir timber	0.55	8,900
Glass	2.5	15,200
Lead	11.0	55,000
Plexiglass or Lucite	1.15	30,800
Steel	7.7	12,700
Gypsum board	0.7	39,000
Oriented strand board[a]	0.6	15,000
Plywood[a]	0.5	16,000

[a] These materials are orthotropic and show very broad, shallow coincidence dips. The values given estimate the frequency at the middle of this dip.

Above the region of coincidence, the sound reduction index depends on frequency and is given by

$$R = R_\perp + 10\log\frac{2\eta f}{\pi f_c}\quad \text{dB}, \tag{12}$$

where η is the loss factor and $R_\perp$ is the sound reduction index for normally incident sound ($\theta = 0$), which is given by

$$R_\perp = 20\log\frac{\pi f M}{\rho_0 c}\quad \text{dB}. \tag{13}$$

Thus, for a constant loss factor, the sound reduction index increases by 9 dB/octave above the critical frequency. The data shown in Fig. 1 for a 150-mm concrete slab increase about 8 dB/octave. In reality, damping in materials is frequency dependent.

The coincidence effect can greatly influence the sound insulation, and where possible, it seems best to position the critical frequency below or above the important frequency range where single-number ratings are calculated. To place f_c at a very low frequency requires a building element with high thickness and stiffness. Heavy walls and floors of brick or concrete usually fulfill this requirement. In any case, it is not usually practical to change the thickness or material properties of such walls. The data for 150-mm concrete in Fig. 2 show that most of the frequency range is above coincidence. The dotted line in the figure is the mass law prediction for a limp layer of the same weight as the concrete.

To place f_c at a high frequency requires a thin and heavy plate of low stiffness. Because of the shape of the reference contours for R_w and STC, unless the coincidence dips at high frequencies are very deep, they do not often reduce the single-number rating.

In some cases, because of the coincidence effect, using a thicker layer of material may not improve R_w or STC. Although mass per unit area is increased, the coincidence dip may be shifted into the frequency range most important for determining the single-number rating. This can be avoided by using two or more layers with the same total thickness. When the layers are attached normally using screws or dabs of glue, the bending stiffness of the double layer is not significantly altered from that of a single layer. Thus the benefits of additional mass per unit area are retained without incurring the loss due to coincidence. The data for the double layer of gypsum board in Fig. 5 give one example of this. The ratings for laminated glass in Fig. 17 give others.

For a random distribution of angles of incidence (diffuse field) the sound reduction index R_{diff} is given by

$$R_{\text{diff}} = R_\perp - 10\log 0.23R_\perp\quad \text{dB}. \tag{14}$$

In practice, sound fields in rooms do not generate an exactly random distribution of angles of incidence on the surface of a building element. There is a lack of waves at grazing incidence, and real panels are not infinite in size. A commonly used approximation is the *field incidence mass law*

$$R = R_\perp - 5 = 20\log Mf - 47\quad \text{dB}. \tag{15}$$

2.8 Orthotropic and Corrugated Materials

Many common building materials are not isotropic, and the bending stiffness varies for different directions along the panels. Plywood and oriented strand board are two such materials that are quite strongly orthotropic. Such materials show a greatly broadened coincidence dip (see Fig. 1). Methods for calculating the sound insulation of such materials are given in Refs. 26 and 27.

Corrugating a sheet of material or adding ribs to it increases the bending stiffness in one direction and creates a new coincidence frequency. This can result in significant reductions in sound insulation, as shown in Fig. 4. Methods for estimating sound transmission through corrugated sheets are given in Refs. 26 and 28.

2.9 Single-Leaf Poured Concrete

To avoid a reduction in sound insulation due to transmission through the pores of materials, porous building elements such as foamed lightweight concrete should be plastered on at least one side. Inhomogeneous elements

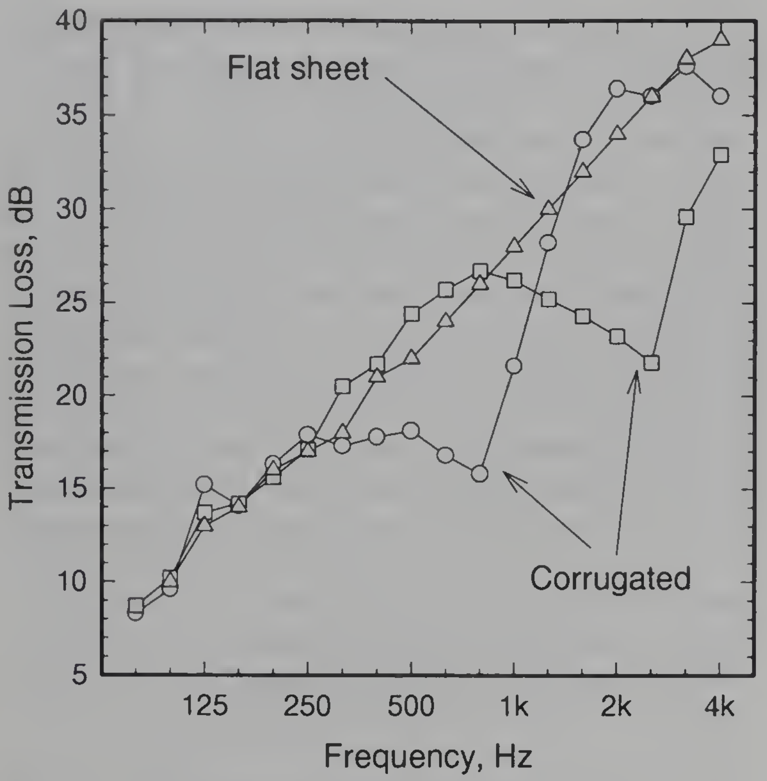


Fig. 4 Sound transmission loss for 0.8-mm steel sheets (6 kg/m^2), flat and with two types of corrugations.

such as floors and exterior masonry with large cavities may provide lower sound insulation because of resonance effects.[29] Small cavities, however, such as in hollow bricks, mostly do not influence sound insulation.

Table 2 gives representative STC values for some common concrete walls or floors measured in laboratories. It will be seen that simple concrete partitions can provide STC ratings from about 45 to 58. To achieve an STC much greater than 58, the weight required is likely to be prohibitive in modern buildings, because the surface mass must be doubled for an increase of 6 dB. The usual solution where high STC ratings are needed is to use multilayer partitions—a central massive wall with one or more leafs attached to each side. This is discussed in Section 2.12.

TABLE 2 Sound Transmission Class for Some Common Concrete Constructions

Concrete Construction	Surface Mass (kg/m^2)	STC
Solid slab		
50 mm	115	43
70 mm	161	47
100 mm	230	52
150 mm	345	55
200 mm	460	58
Hollow-core slab		
150 mm	220	48
200 mm	278	50
250 mm	312	50
Precast double T 355 mm deep with 50 mm topping	366	54

2.10 Single-Leaf Masonry Walls

In the thicknesses commonly used in buildings, hollow-core concrete block provides a few decibels less sound insulation than a solid concrete wall or brick wall of the same weight per unit area, provided all cracks, openings, and voids are properly sealed.

Solid and hollow-core concrete blocks are manufactured in a variety of thicknesses, core sizes, aggregates, and densities. Representative values of the STC for block walls are given in Table 3. These values are only valid if the wall surfaces are properly sealed and the mortar joints are sound.

The sound insulation provided by a hollow-core block depends not only on the density of the block material but also on its porosity. The more porous the block, the more sound will leak through the block structure and the greater will be the improvement if the surfaces of the block are sealed (e.g., by concrete paint, epoxy paint, or a skim coat of plaster). It is enough to apply the sealant on one face of the block only. In fact, when gypsum board is

TABLE 3 Sound Transmission Class Ratings for Standard Hollow Normal and Lightweight Block Walls Sealed on at Least One Side

Nominal Thickness (mm)	Lightweight			Normal Weight		
	kg/block	kg/m^2	STC	kg/block	kg/m^2	STC
100	8	105	43	10	130	44
150	10	130	44	15	190	46
200	14	180	46	18	225	48
250	17	215	47	21	260	49
300	20	250	49	25	310	51

to be added on resilient supports to finish the wall, this can be an advantage (see Section 2.12). Improvements of 5–10 STC points are not uncommon after sealing. A layer of gypsum board attached directly with screws or dabs of glue to the block surface is *not* an effective seal because it is still able to vibrate as a separate layer. Even when gypsum board on studs or resilient supports is to be used to finish a block or concrete wall, the wall should be sealed. Covering a flaw with gypsum board does not eliminate the detrimental effects of the flaw.

The porosity of acoustic materials is characterized by the airflow resistivity. This quantity is calculated from the volume velocity of air flowing through a specimen and the air pressure drop across it.[30,31] Blocks with airflow resistivities greater than about 2×10^5 Pa · s/m^2 show no improvement after sealing. Normal-weight hollow blocks and solid blocks are not likely to need sealing. Lightweight blocks usually do. In the absence of information on block airflow resistivity, to ensure the maximum sound reduction from a block or masonry wall, it is safer to seal the wall properly on one side. In any case, all the mortar joints must be properly finished and free from leakage.

Table 4 gives approximate sound insulation values for single-leafed building elements. These values apply when flanking transmission exists that is typical of that for construction having a mean mass per unit area of 300 kg/m^2.

2.11 Double-Leaf Building Elements

When high sound insulation is required without using heavy single-leaf constructions, the best approach is to construct a wall consisting of two solid leaves with an airspace or resilient layer between them. When properly designed, such a construction provides sound insulation greater than that for a single layer of the same mass per unit area and thickness. The total sound insulation of a double-leafed building element is, however, lower than the sum of the sound insulation of each leaf alone. This is because of the coupling through the air in the cavity and through any physical ties between the layers.

Figure 5 shows the sound reduction index through 16-mm gypsum board for four cases: (1) installed as a single layer, (2) installed as two layers screwed together, (3) installed as two layers with a cavity between them, and (4) installed as two layers with a cavity filled with glass fiber batts. This figure shows typical behavior for double-layer walls where the studs are resilient or there are independent studs for each layer of gypsum board. At the frequency marked f_r there is a dip in the transmission loss curve that reduces the sound insulation below

TABLE 4 Sound Insulation of One-Leafed Building Elements with Mass per Unit Area *M* and Thickness *d* with Flanking Typical of Construction with Mass per Unit Area of 300 kg/m^2

	R' (dB) at f (Hz)								
Material	125	250	500	1000	2000	4000	M (kg/m^2)	d (mm)	R_w^1
Hollow brick with 15 mm plaster on both sides	36	42	48	55	56	59	350	270	52
	34	40	40	46	51	56	210	145	46
Brick with 15 mm plaster on both sides	40	46	51	54	59	62	480	270	55
	36	42	42	48	53	58	260	145	48
	31	39	39	37	49	53	170	105	42
	31	37	37	34	47	52	150	95	37
Dense concrete, without plaster	39	43	50	55	62	66	430	190	54
	38	42	47	54	61	64	350	150	52
	34	38	48	53	61	63	300	120	50
	30	33	37	44	51	59	170	70	42
	31	29	27	36	43	48	95	40	37
Lightweight concrete with 10 mm plaster on both sides	30	37	42	50	56	60	190	250	46
	31	32	37	45	50	56	130	220	42
	31	32	34	43	50	55	100	170	40
	28	32	30	36	46	55	85	110	36
	25	29	29	28	38	44	60	70	32
Gypsum block, without plaster	39	40	41	50	55	58	315	260	47
	31	33	34	42	50	60	105	100	41
	32	30	26	34	43	49	85	70	33

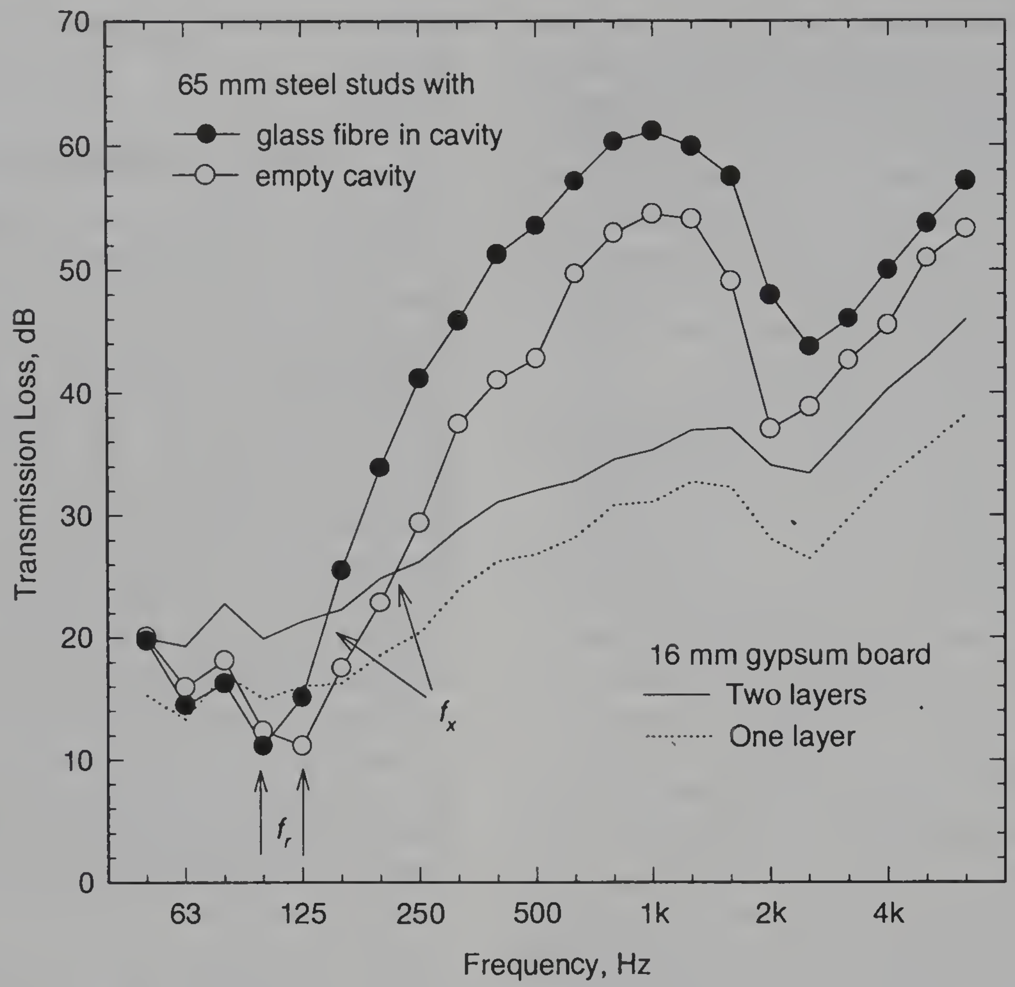


Fig. 5 Sound reduction index of double- and single-leafed gypsum board walls. The 25-gauge (0.5-mm) steel studs in the double wall are 38 × 65 mm deep. The sound-absorbing material is glass fiber with a thickness of 65 mm and a flow resistivity of 3600 N · s/m^2.

that for the two layers in contact and even below that for the single layer. The transmission loss of the double-leafed walls becomes equal to that for two leaves in contact below the resonant frequency at the lowest frequency measured. Above the resonance, the sound insulation rises steeply, crosses the curve for the two layers in contact, and increases for several bands before following a more gradual increase and then turning down into the coincidence dip.

For an air-filled cavity, the mass–air–mass resonance frequency f_r can be calculated from

$$f_r = \frac{1}{2\pi}\sqrt{\frac{\rho_0 c^2}{d M_{\text{eff}}}}, \qquad (16)$$

where d is the distance between the inner surfaces (in metres) and M_{eff} is the effective mass, defined as

$$\frac{1}{M_{\text{eff}}} = \frac{1}{M_1} + \frac{1}{M_2}, \qquad (17)$$

M_1, M_2 being the masses per unit area of the two leaves (in kilograms per square metre).

For a given total mass, this frequency has its minimum value when both masses are equal. Adding sound-absorbing material to the cavity lowers the frequency at which the mass–air–mass resonance occurs. In the cavity, the propagation constant will be quite different from that in air; so the effective value of $\rho_0 c^2$ changes. The shift in the mass–air–mass resonance depends on the flow resistivity and thickness of the sound-absorbing material. As an approximate guide, when the wall cavity contains sound-absorbing material, multiply the frequency calculated from Eq. (16) by a factor of 0.7. Losses in the sound-absorbing material will influence the depth of the resonance.

Prediction of sound transmission through double-layer walls has been studied by several authors.[32–34] Two frequencies are important in making predictions: the mass–air–mass resonance and the first cross-cavity resonance f_1, which is given by

$$f_1 = \frac{c}{2\pi d} \quad \text{Hz}. \qquad (18)$$

Higher cross-cavity resonances are integer multiples of this frequency. Another advantage of having sound-absorbing material in the cavity is that it damps resonances at f_1 and higher cavity modes.

When the two panels are completely isolated from each other, the sound reduction index can be predicted approximately by

$$R = \begin{cases} 20\log[(m_1 + m_2)f] + 47, & f < 2f_r/3, \\ R_1 + R_2 + 20\log fd - 29, & f_r < f < f_1, \\ R_1 + R_2 + 6, & f > f_1, \end{cases} \qquad (19)$$

where R_1, R_2 are the sound reduction indexes for each layer of the double wall measured or calculated separately. Thus, high sound insulation can be attained by selecting a high mass per unit area of the leaves and a large cavity depth to ensure a low mass–air–mass resonance frequency, for example, below 100 Hz.

Procedures have been given for calculating sound reduction indices when the panels are connected along a line or at points,[32] but no reduction for the mass–air–mass resonance is considered. Modified procedures to deal with resilient line connections (light gauge metal studs) and the mass–air–mass resonance have also been suggested.[34] The formulas in Eq. (19) combine the ideas from Refs. 32 and 34 and give an approximate way of dealing with the low-frequency resonances that invariably limit the sound insulation of lightweight double-panel walls.

In practice, it is very difficult to construct a double-panel wall where the two panels are not connected in some way, and the sound reductions predicted by Eq. (19) are seldom achieved. Except in exceptional cases, there is often a solid connection around the edges. The more rigid the connections between the panels, the lower the sound reduction. Wood studs joining the two panels provide line connections between them. Such solid connections should be avoided where possible. Resilient metal channels on one or both sides of solid wood or stiff metal studs reduce the transmission of vibration from one panel to the other and give better sound insulation but still not as good as completely isolated studs. If connections must be solid, then it is better to connect only at points rather than along a line. In all cases, connections to panels should be minimized; the closer the spacing of studs or resilient metal channels, the lower the sound reduction.[35]

Typical methods for providing mechanical isolation between double-wall layers are illustrated in Fig. 6. Non-load-bearing steel studs [typically made from 25-gauge (0.5-mm) sheet steel] are usually flexible enough to reduce substantially the vibration transfer between the layers on each side.

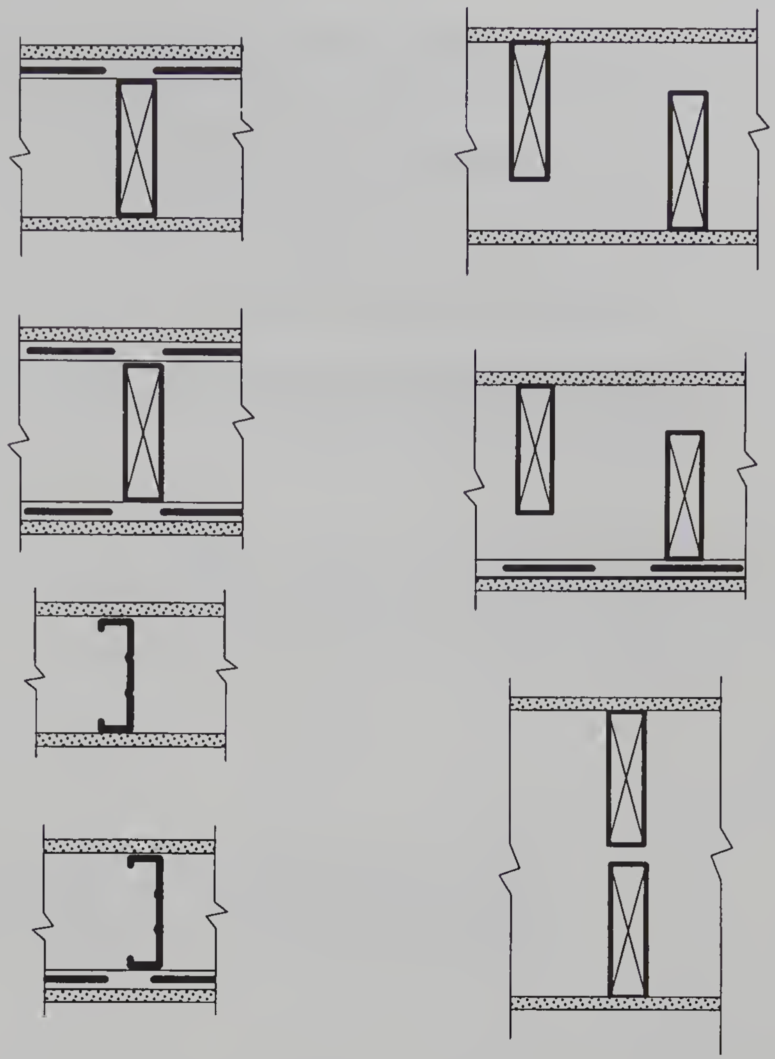


Fig. 6 Methods for providing mechanical isolation between partition layers: staggered wood studs with and without resilient metal channels and sharing a common sole and header plate, separate rows of wood or steel studs with independent sole and header plates, a single row of wood or load-bearing studs with resilient metal channels, and a single row of light-gauge metal studs.

The more sound-absorbing material in the cavity, the higher the sound insulation, but the greatest increases occur for the first 25% or so of the cavity depth.[35,36,37] Incremental increases after the cavity is about 75% full are usually small. Figure 7 shows an example of measurement results showing this effect. The absorbent material should have[38] an airflow resistivity of more than about 5 kPa · s/m^2, a value satisfied by most common fibrous materials. Higher values of resistivity do not significantly increase the STC or R_w, although there are significant decreases in sound transmission at higher frequencies.

If the cavity is tightly filled with a fairly dense absorptive material, there is some risk that this will cause connections of high stiffness between the leaves and lower the sound insulation; sound will propagate through the body of the sound-absorbing material. Overfilling the cavity should be avoided.

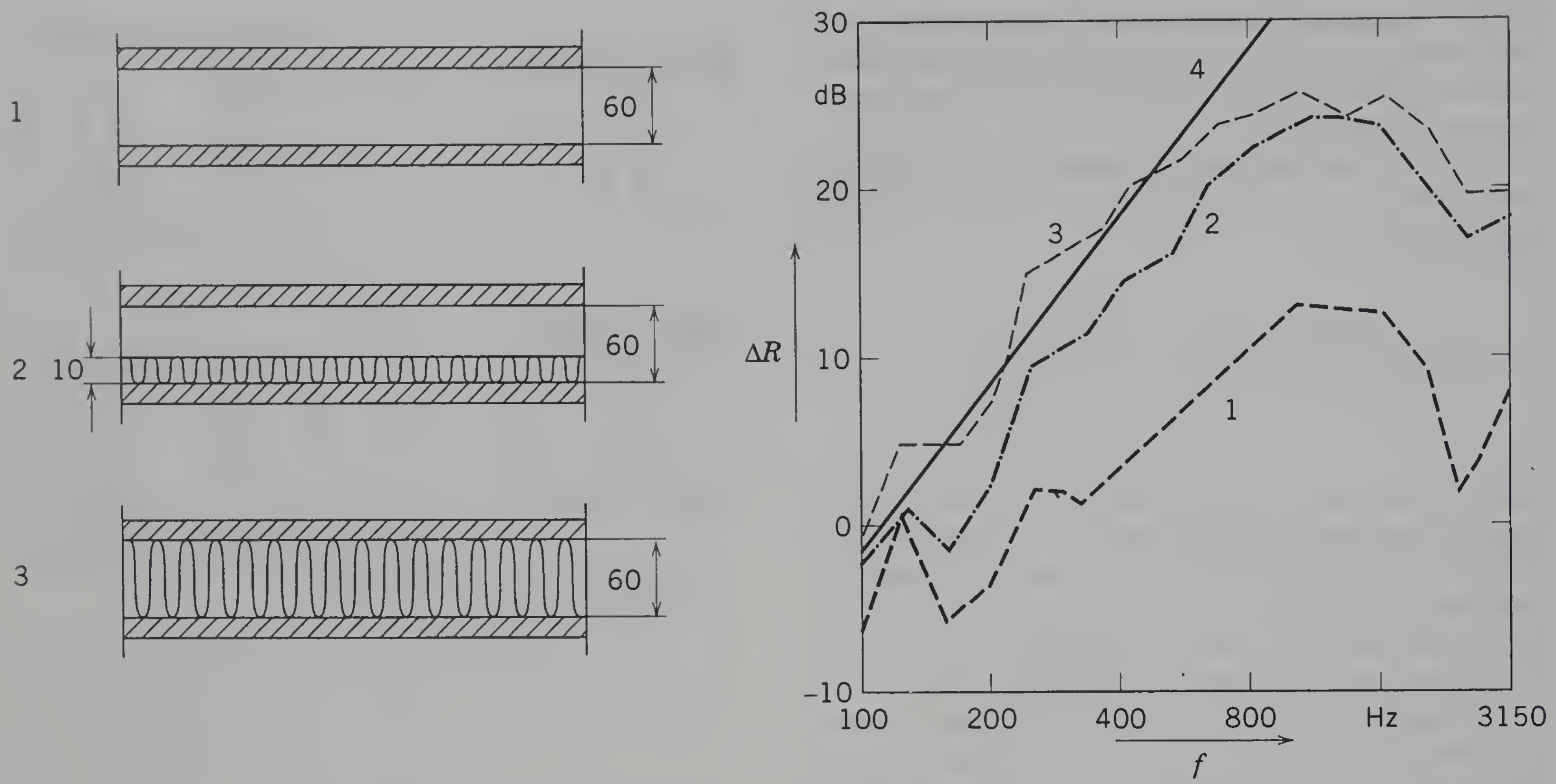


Fig. 7 Increase of sound reduction index ΔR by double-leafed walls of 12.5-mm gypsum board (cavity width 60 mm) relative to a single-leaf wall of the same mass per unit area: (1) without absorbent lining; (2) 10 mm mineral wool; (3) cavity filled with mineral wool; (4) calculated values.

The same acoustical and construction principles have to be taken into account when using a suspended ceiling to improve the sound insulation of a floor. The suspended ceiling will increase both airborne and impact sound insulation (see Sections 3.3 and 3.6). Materials like wood-wool slabs, plastering on expanded metal. gypsum board, and chipboard are in use for suspended ceilings. Best results are obtained with resilient suspensions or independent ceiling joists.

Double-Leaf Gypsum Board Walls Table 5 gives measured STC values for some walls constructed from wood or steel studs and 13- or 16-mm gypsum board.[39] The surface mass of the fire-rated gypsum board was respectively about 10 and 11 kg/m^2. Other lighter types of board were used in the study and gave correspondingly lower results. In each case, the cavity was about three-fourths to completely full of glass fiber batts. The type of sound-absorbing material has only a small effect on the single-number rating.[40] Other constructions with similar structural connections and cavity depth should give similar values of sound insulation.

In determining the sound reduction of cavity walls, the two major factors are the total mass of the layers and the depth of the cavity. For typical gypsum board cavity walls it has been found that doubling the total mass or the cavity depth increases the STC or R_w by about 10 points.[40]

Adding a single, solid, nonporous layer to one set of studs in the middle of a double stud wall with sound-absorbing material in both stud cavities reduces the sound insulation by about three to four points. A second layer added internally to the other set of studs would make the sound insulation even worse. In practical situations, for a given total weight, single-cavity walls perform better overall than walls with multiple cavities.

A casual examination of the cross section for a staggered wood stud wall suggests that there will be no solid connection between the layers of gypsum wallboard on each side. This conclusion overlooks the connection through the sole and header plates. Sound transmission around the periphery of such walls reduces the sound reduction for the wall. Table 5 shows that adding resilient metal channels to one side of the studs reduces this peripheral transmission and increases STC ratings. While the channels increase the depth of the cavity, this is not enough to account for the increase in STC over the case with no channels. The improvement comes from reducing the transmission through the plates around the edges of the wall.

For stud walls, even for a double stud wall, the sound reduction is influenced by the spacing of the studs and

TABLE 5 Approximate STC Ratings for Walls with 13- and 16-mm Gypsum Board on Both Faces

	Gypsum Board 16 mm			Gypsum Board 13 mm		
	1|1	1|2	2|2	1|1	1|2	2|2
38 × 89 mm wood studs, 400 mm oc	34 (37)	36 (40)		34 (39)	37 (41)	38 (43)
38 × 89 mm wood studs, 400 mm oc, resilient metal channels on one side, 600 mm oc	46 (47)	53 (52)	59 (57)	46 (48)	51 (52)	57 (57)
Staggered 38 × 89 mm wood studs	47 (48)	52 (52)	56 (56)	47 (49)	50 (52)	55 (56)
Staggered 38 × 89 mm wood studs, resilient metal channels on one side, 600 mm oc	51 (52)	56 (57)	62 (64)	49 (52)	54 (57)	60 (62)
Double row of 38 × 89 mm wood studs, 400 mm oc	58 (58)	62 (64)	67 (69)	58 (58)	62 (63)	66 (68)
Double row of 38 × 89 mm wood studs, 600 mm oc	59 (60)	64 (65)	69 (69)	57 (58)		
65-mm steel studs, 25 ga, 400 mm oc	39 (42)	45 (47)	52 (52)	36 (41)	42 (46)	48 (50)
65-mm steel studs, 25 ga, 600 mm oc	44 (44)	51 (49)	55 (53)	45 (44)	51 (49)	55 (54)
90-mm steel studs, 25 ga, 400 mm oc	49 (47)	52 (50)	56 (55)	46 (45)	51 (50)	55 (53)
90-mm steel studs, 25 ga, 600 mm oc	49 (48)	54 (53)	57 (55)	48 (47)	52 (52)	55 (54)
150-mm steel studs, 25 ga, 600 mm oc	51 (51)			52 (52)		
90-mm steel studs, ≤20 ga, 400 mm oc	49 (49)			48 (49)	53 (53)	59 (59)

Note: Numbers in parentheses are R_w ratings. The notation 1|1, 1|2, 2|2 indicates the number of layers of gypsum board on each face, oc = on center, spacing between studs; ga = gauge.

resilient channels that support the gypsum board. Within practical limits, it is best if these are as far apart as possible. Thus a 600-mm spacing gives better results than 400-mm spacing. For double wood studs the increase is about two points.

When resilient metal channels are used to support gypsum board on wood studs and the construction is unbalanced (two layers of gypsum wallboard on one side and one on the other), it does not matter which side the channels are on; the STC rating is the same. However, the assembly with two layers of gypsum wallboard attached to the resilient metal channels gives a better fire resistance rating.

In concrete buildings, there is invariably a certain amount of flanking sound transmision that reduces the overall sound insulation. This reduction can be calculated as discussed in Section 2.15. Figure 8 gives some typical values of sound insulation for double-leafed lightweight walls in situ. When joist construction is used exclusively in a building, it is very much more difficult to calculate or predict flanking, and the data in Fig. 8 are not applicable to such buildings.

Double-Leaf Masonry or Concrete Walls Two-leaf masonry or concrete walls can, in principle, provide very high sound insulation because they comprise two very heavy layers separated by an airspace. The maximum insulation attained is usually limited, however, by the practical difficulties of constructing two leafs that are structurally isolated. Depending on the height of a block wall, metal ties are required for structural reasons or to meet applicable building codes. Solid metal ties transmit sound energy from one leaf to the other, but special ties, designed to reduce sound transmission, can minimize the detrimental effects of mechanical ties.

Perhaps more important is transmission of energy along the floor and ceiling, along walls abutting the periphery of the cavity wall, and through other parts

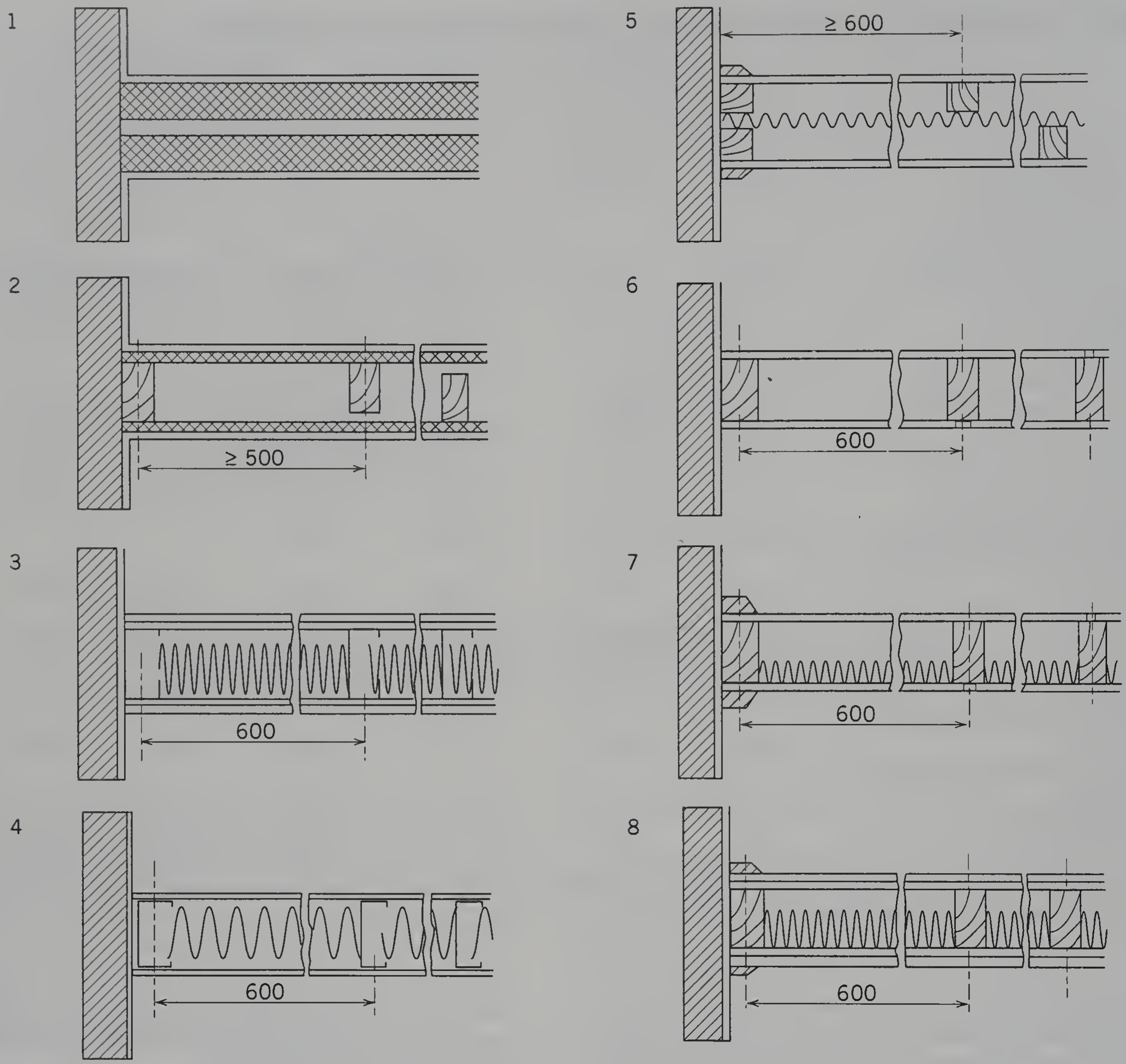


Fig. 8 Sound insulation for double-leafed lightweight walls in concrete buildings. (1) Wood–wool (50 mm), plastered; cavity (30 mm); $R'_w = 52$ dB. (2) Wood–wool (25 mm), plastered double-studs; cavity (100 mm); $R'_w = 52$ dB. (3) Plasterboard (2 × 12.5 mm); cavity (100 mm); absorptive quilt; sheetmetal studs; $R'_w = 52$ dB. (4) Like 3, but plasterboard (1 × 12.5 mm); $R'_w = 47$ dB. (5) Like 2, but plasterboard (12.5 mm) and added absorptive quilt (40 mm); $R'_w = 47$ dB. (6) Plasterboard (12.5 mm); wooden studs; cavity (100 mm); $R'_w = 37$ dB. (7) Like 6, but added absorptive quilt (40 mm); $R'_w = 42$ dB. (8) Like 7, but plasterboard (2 × 12.5 mm); $R'_w = 48$ dB.

of the structure. These *flanking paths* bypass the cavity wall and reduce its apparent sound insulation. This kind of flanking is illustrated in Fig. 9 and discussed further in Section 2.15. Physical breaks in the floor, ceiling, and abutting walls are needed to reduce transmission along such paths. However, even if gaps are included in the design, during construction the airspace between leafs may become bridged by mortar droppings or rubble that will reduce the sound insulation. Such defects are concealed and impossible to correct after the wall is complete. In practical installations careful design of the whole system, good supervision, and care during con-

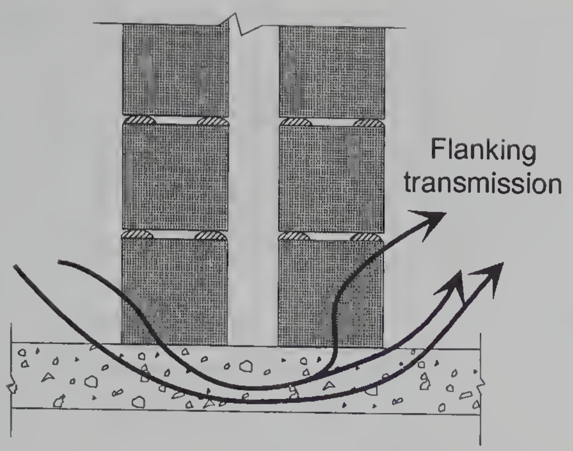


Fig. 9 Flanking paths under a cavity block (or concrete) wall resting on a continuous slab of concrete. Any solid connections between the block walls will also reduce the sound insulation.

struction are needed for two-leaf block walls to achieve their potential.

Figure 10 shows laboratory sound transmission loss results for two-leaf walls formed from two isolated 90-mm block layers with a cavity of 125 mm between them. The figure also shows results for a double-layer wall formed from 90- and 190-mm blocks with a cavity of 165 mm between them. The cavity between the blocks had a 65-mm-thick layer of glass fiber, 50-mm-thick layer of polystyrene foam, and only air in the cavity. The benefits of the glass fiber batts are clear. Table 6 gives sound insulation values for double-leafed heavy walls assuming typical conditions of flanking transmission. Although sound-absorbing material is very effective inside lightweight constructions or in heavy cavity walls in the laboratory, because of flanking, it has almost no effect on heavy concrete or masonry walls in field situations.

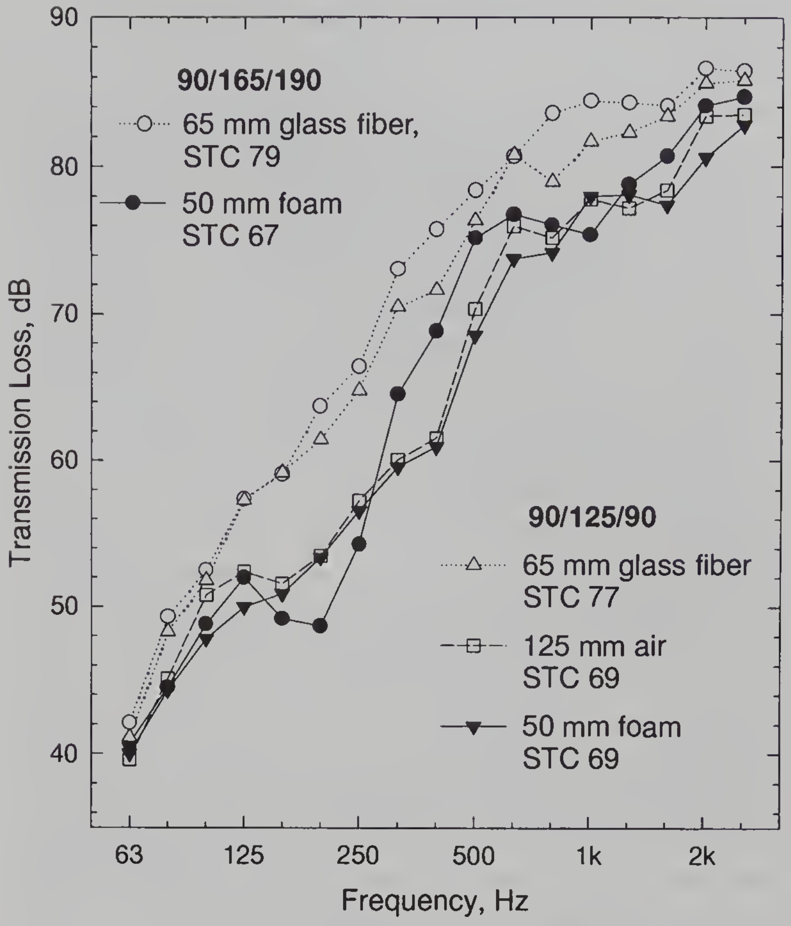


Fig. 10 Double-layer concrete block walls measured in a laboratory. One wall was supported on the edge of one of the reverberation rooms and the other on an independently supported frame. One wall was formed from two isolated 90-mm block layers with a cavity of 125 mm between them. The figure also shows results for a double-layer wall formed from 90- and 190-mm blocks with a cavity of 165 mm between them. The cavity between the blocks had a 65-mm-thick layer of glass fiber, 50-mm-thick layer of polystyrene foam, and only air in the cavity.

TABLE 6 Apparent Sound Reduction Index of Double-leafed Building Elements

Construction	1st layer (mm)	Air gap (mm)	2nd layer (mm)	Sound Reduction Index by Frequency 125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	M (kg/m^2)	Total width (mm)	R'_w (dB)
Dense concrete, without absorbent material in the cavity	40	25	70	33	38	43	50	57	55	275	135	49
	70	10	70	43	44	50	54	55	60	340	150	53
	40	50	70	35	42	45	53	58	60	275	160	50
	70	50	70	44	42	48	54	59	58	340	190	52
	40	100	70	44	42	47	55	58	62	275	210	52
	70	100	70	43	41	48	54	59	65	340	240	53
Lightweight concrete, mineral wool in the cavity (50 mm)	70	110	120	42	44	46	48	53	60	175	300	51
	70	160	70	38	41	42	44	52	60	135	300	47
	70	50	70	37	43	41	44	55	63	135	190	47
	115	80	115	45	42	46	59	56	64	190	310	52
Gypsum, without absorbent material in the cavity	70	60	70	32	40	39	45	53	64	170	200	46
	80	60	80	38	42	41	46	54	62	210	220	47
Gypsum, mineral wool in the cavity (30 mm)	60	30	60	39	40	40	48	55	64	100	150	48
	70	60	70	35	40	41	46	56	63	160	200	47
	80	30	80	36	41	39	43	52	67	170	190	46

2.12 Block or Concrete Walls with Added Gypsum Board

The highest sound insulation that can be attained in practice from a single-leaf wall is limited by the mass. To get STC or R'_w ratings much greater than about 58, requires surface weights of 500 kg/m^2 or more. This sound insulation may not be enough or the weight may be unacceptable. In practice, the sound insulation of a heavy wall is often improved by adding layers of gypsum board, or something similar, supported independently or on resilient mounts to form a cavity construction.

Gypsum board is usually added on both sides of the wall, so the construction becomes a triple wall. The important parameters—mass–air–mass resonance, separate studs or resilient furring to avoid direct transmission between the gypsum board and the blocks, and the use of sound-absorbing material in the cavity—are doubly important. Reductions in sound insulation due to the mass–air–mass resonance are doubled but, conversely, improvements are doubled too.

Experiments with 190-mm-thick concrete blocks[41,42] showed the importance of avoiding small cavity depths that result in the mass–air–mass resonance being within or near the range used in calculating the STC or R_w (Fig. 11). Adding sound-absorbing material to the cavity clearly lowered the mass–air–mass resonance frequency (Fig. 12). Adding gypsum board on both sides of the blocks caused the mass–air–mass resonance to be deeper than it was when the gypsum board was applied on just one side (Fig. 13). When the cavity depth was only 13 mm, the STC was reduced by one point, but the detrimental effects on sound insulation below 160 Hz were much more serious.

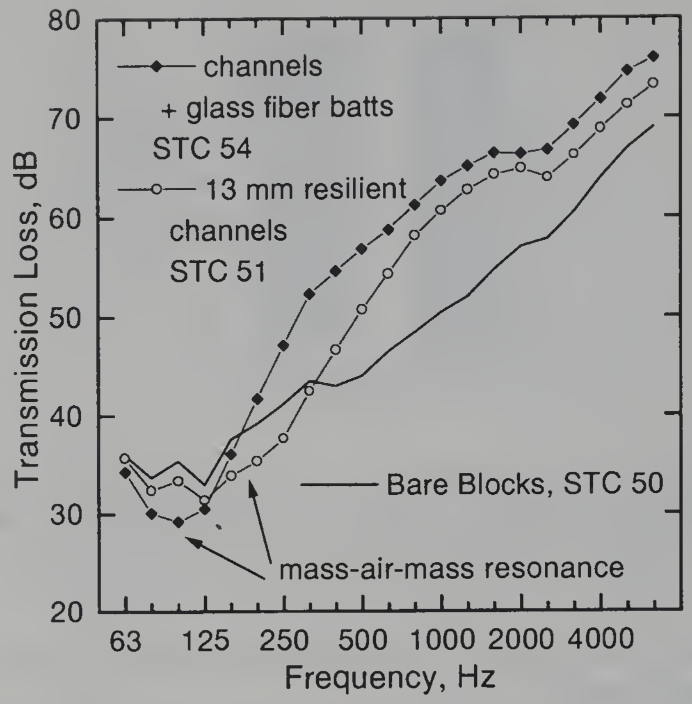


Fig. 12 Sound transmission loss for a 190-mm concrete block wall with 16-mm gypsum board attached on 13-mm resilient metal channels on one side of the wall with and without sound-absorbing material in the cavity.

Some concrete blocks are sufficiently porous that the cavity formed behind the gypsum board has an effec-

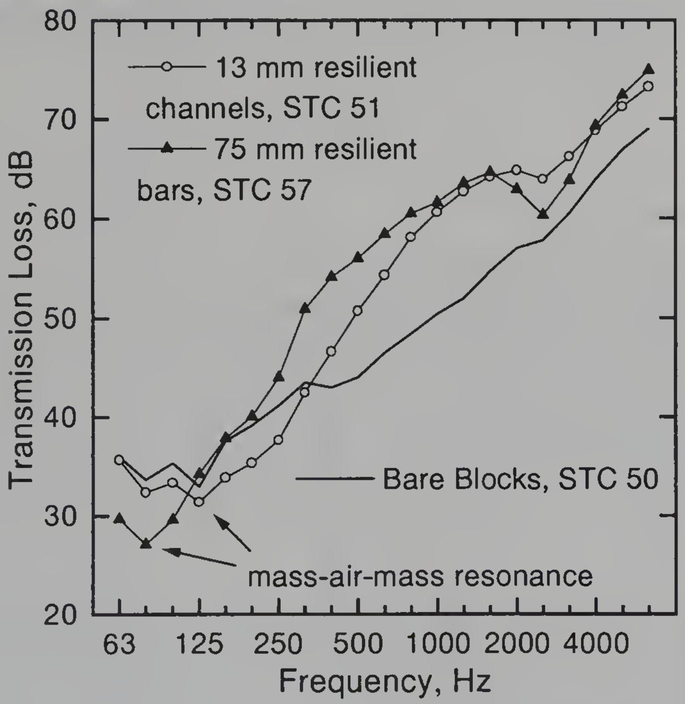


Fig. 11 Sound transmission loss through a 190-mm concrete block wall with 16-mm gypsum board attached on 13-mm resilient metal channels and 75-mm-deep resilient metal Z-bars.

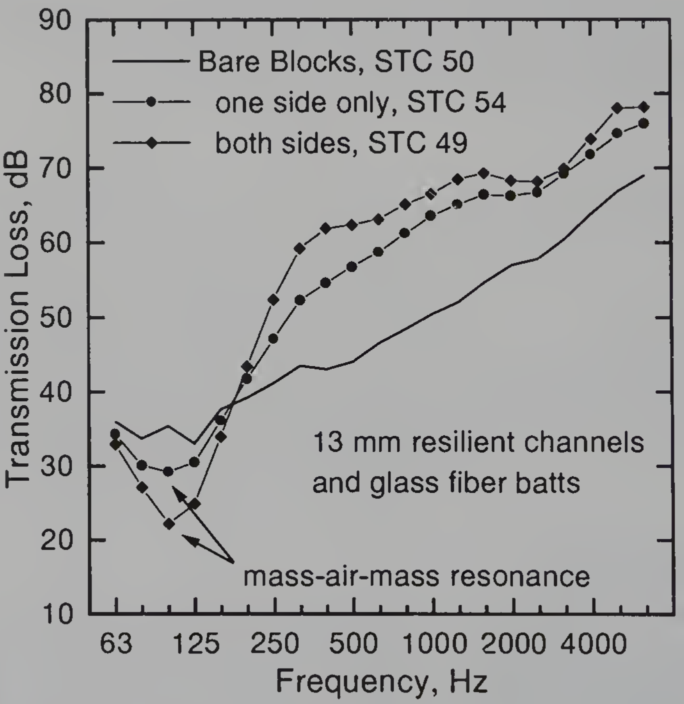


Fig. 13 Sound transmission loss for a 190-mm concrete block wall with 16-mm gypsum board attached on 13-mm resilient metal channels to one side and to both sides of the wall with sound-absorbing material in the cavity.

tive depth that is greater than the distance from the surface of the blocks to the rear face of the gypsum board; the air in the cavity penetrates the block.[43] This lowers the mass–air–mass resonance and increases sound insulation, but the blocks do not provide the full sound insulation expected from their mass. When the blocks are extremely porous, the sound insulation obtained is determined almost completely by the outer layers of gypsum board. As pointed out above (Section 2.10), the reduction in sound insulation can be prevented by sealing one face of the blocks with plaster or latex point. On the sealed side of the blocks, of course, there is no longer an increase in effective cavity depth because of porosity, but on the unsealed side there can be considerable benefit (see Fig. 14).

The measured improvement in STC rating that resulted from the attachment of 16-mm gypsum board in a variety of ways to a concrete block wall is given in Table 7. The STC ratings for complete wall systems can be calculated by adding the improvements given in Table 7 to the STC rating of the masonry walls from Table 3 or the concrete walls from Table 2. Note again that when the cavity behind the gypsum board is too small, the STC is actually lowered relative to the bare-block case. Note also the very large increases in STC when the cavity is large and filled with sound-absorbing material (see Fig. 14).

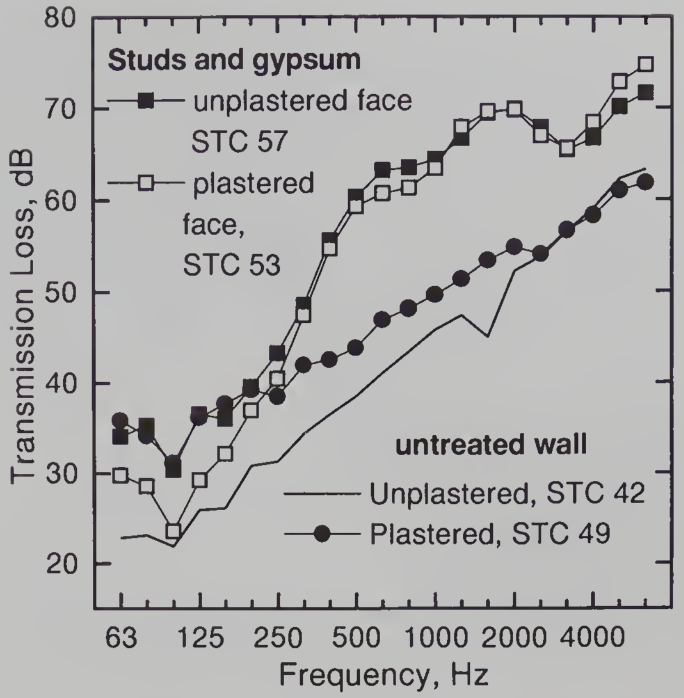


Fig. 14 Effect of plastering one face of a 190-mm lightweight block and of adding 16-mm gypsum board and 40-mm steel studs on each face. Note the reduced effect of the mass–air–mass resonance when gypsum board is applied on the porous face.

2.13 Double Walls Containing Resonators

Some research has been done to find a means of counteracting the mass–air–mass resonance to improve the low-frequency sound insulation of double walls by including Helmholtz resonators within the double-wall cavity. Data presented for double-panel constructions where there were Helmholtz resonators around the edge of the cavity[44,45] showed that a correctly chosen Helmholtz frequency improves sound insulation around the mass–air–mass resonance frequency.

Certain types of concrete blocks have slots in one face opening into the internal block cavities and are designed to act as Helmholtz resonators. The blocks absorb sound quite effectively at the resonance. These resonator blocks offer a way of counteracting the deleterious effects of the mass–air–mass resonance in block walls. By constructing the core wall so the slots are exposed equally on both faces of the wall, each layer of gypsum board will have a resonant structure behind it. In Fig. 15 the sound transmission losses for normal-weight and resonant blocks

TABLE 7 Increase in STC Ratings with Single Layer of 16-mm Gypsum Board Added to One or Both Sides of Concrete Block or Poured Concrete Wall with and without Fiberglass Batts Filling Cavity between Gypsum Board and Concrete

	Without Fiberglass in Cavity		With Fiberglass in Cavity	
Gypsum Board Attachment	One Side	Both Sides	One Side	Both Sides
Directly on concrete block	+0	−1		−1
On 13-mm resilient steel channels	+2	−1	+4	+14
On 50-mm resilient steel furring	+2	+2	+9	+22
On 65-mm steel studs	+8	+7	+10	
On 75-mm resilient steel furring	+7		+11	

Note: The same values may be used for 13-mm gypsum board without serious error.

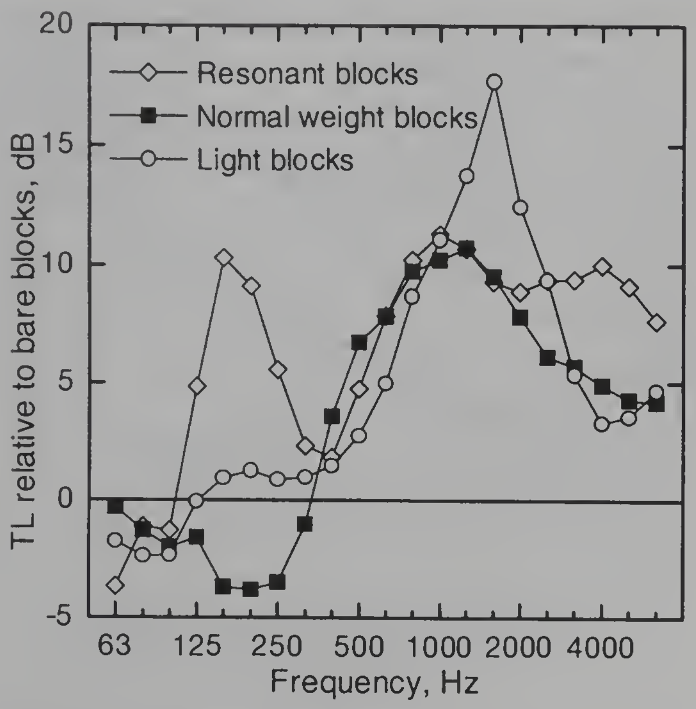


Fig. 15 Transmission loss relative to untreated concrete blocks for three kinds of blocks with gypsum board and 13-mm resilient metal channels on one face only.

with gypsum board supported on resilient metal channels are compared.[46] To eliminate the differences between the basic blocks, the transmission losses for the composite walls are shown as differences from those for the basic blocks in each case. The differences in the curves around 200 Hz are quite striking. For the normal-weight blocks, the mass–air–mass resonance causes a reduction in sound transmission loss. The resonant blocks show a strong absorptive resonance at 160 Hz, the Helmholtz resonance for these blocks, that counteracts the mass–air–mass resonance and leads to increased STC and R_w. The cost of a construction using such blocks might be justifiable in some cases.

2.14 Doors and Windows

Since doors and windows must be operable and therefore light, the sound insulation available from mass alone is limited. High insulation can be obtained, however, by the use of double-leafed elements, that is, double doors or double windows. As with any double-leafed system, the

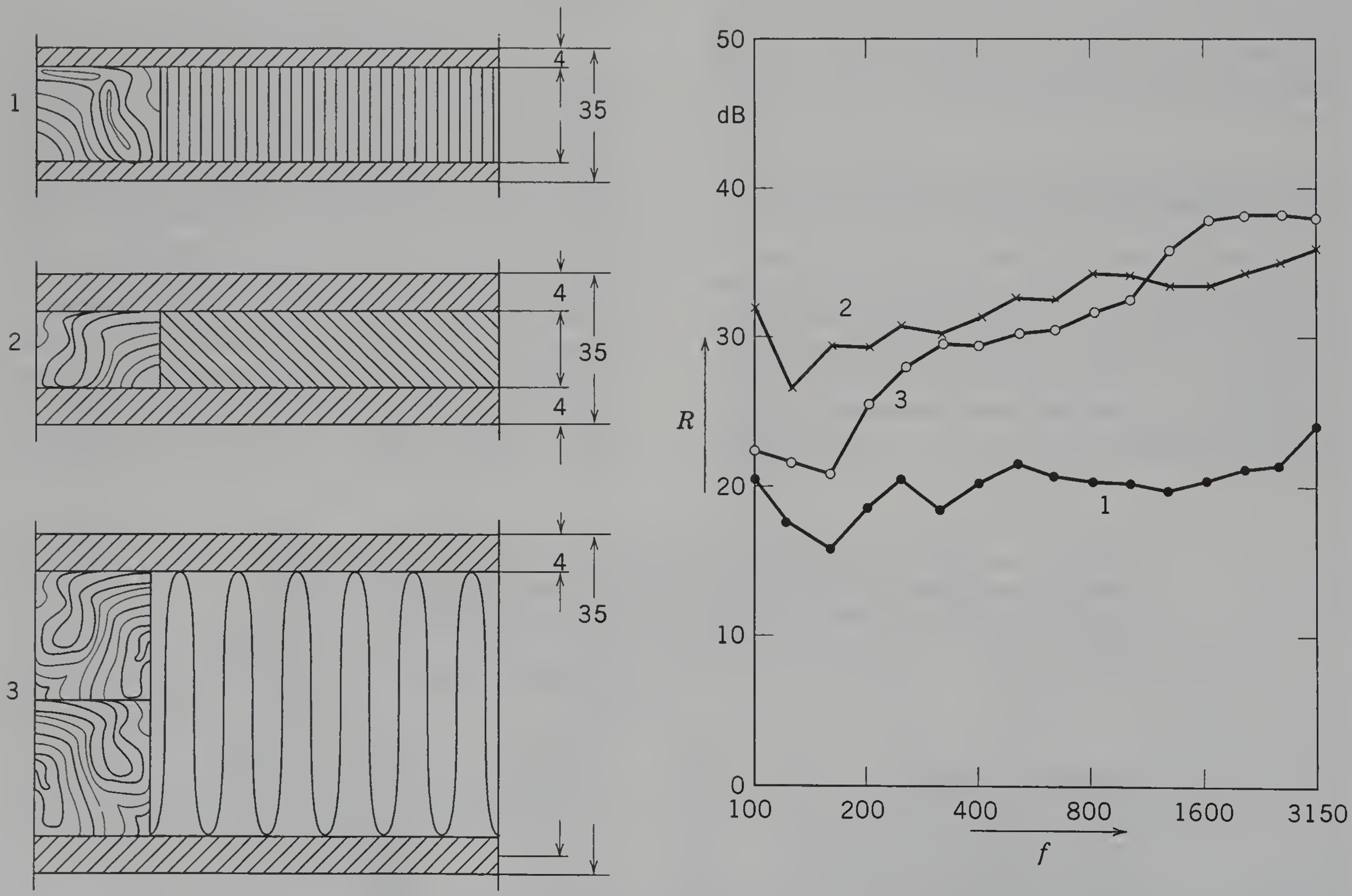


Fig. 16 Sound reduction index R of doors vs. frequency. (1) Hardboard faced with honeycomb skeleton: $M = 11\ \mathrm{kg/m^2}$; $R_w = 12$ dB. (2) Gypsum board between wood chipboard: $M = 42\ \mathrm{kg/m^2}$; $R_w = 35$ dB. (3) Double-leafed door with wood chipboard; mineral wool lining: $M = 23\ \mathrm{kg/m^2}$; $R_w = 32$ dB.

width of the airspace should be as large as possible. In the case of doors the distance should be increased to form a lobby or corridor whenever possible. An absorbent lining will improve sound insulation.

In Fig. 16 the sound reduction index of three typical door leaves sealed at the edges is shown. For lightweight doors made, for example, of hardboard-faced flush skeleton ($M \approx 10$ kg/m^2), a weighted sound insulation index of 20–25 dB cannot be exceeded. An increase in the door weight ($M \approx 45$ kg/m^2), attained by using cored panels or two or more solid layers such as gypsum board or wood chipboard, improves sound insulation to about 35–40 dB. Higher sound insulation can be gained with well-designed double-leafed panels (see Section 2.11). The use of double doors may result in an improvement of about 10 dB, provided there is a sufficient distance (> 100 mm) between them.[47]

Proper seals and weather stripping are essential. A decrease of sound insulation of about 10–20 dB can be expected if the doors or windows have poorly fitting seals. Figure 17 shows some examples of single- or double-sealed door openings. Of course, the higher sound insulation of the panel, the better the sealing must be. If a threshold at the foot of the door must be avoided, sealing can be particularly difficult. Figure 18 gives some effective solutions to meet sound insulation requirements.

The sound insulation of single-glazed windows is influenced by coincidence with the critical frequency f_c, which may be calculated from the information given in Table 1. The sound insulation does not increase with increasing thickness as rapidly as might be expected from mass law, but some improvement can be obtained by using laminated glass (security glass). This consists of thin glass panes with an intervening plastic layer that minimizes the increase of stiffness due to increasing thickness. Figure 19 gives the weighted sound reduction index obtainable by single-glazed windows both with normal and laminated panes. The values in the figure may be calculated using

$$R_w = 11.7 \log d + 23.1 \quad \text{for solid panes} \qquad (20)$$

or

$$R_w = 17.9 \log d + 19.2, \qquad d \geq 4 \quad \text{for laminated panes,} \qquad (21)$$

where d is the total thickness of the pane in millimetres.

To obtain an improvement in sound insulation by using double-glazed windows requires a cavity depth of about 20 mm or more. An airspace smaller than 100 mm is inadvisable if a weighted sound reduction index of more than about 40 dB is to be obtained. In Fig. 20 the approximate weighted sound reduction index of double-glazed window systems is shown for different total thicknesses of both glass panes. Higher values can be obtained by using panes of different thicknesses to reduce the influence of coincidence. A ratio of about 1 : 2 in thickness is recommended. As much absorption as possible can be provided in the cavity by lining the perimeter (reveals, head, and sill if possible) with an efficient sound-absorbing material. This gives most benefit at high frequencies. Further improvements can be made using resonant cavities around the perimeter.[44,45] The data in the figure may be calculated using $R_w = 25(1 + \log d)$.

With double windows it is even more important than with single windows to ensure proper sealing. Just as with door openings, resilient materials of high durability must be used. To have the seals fit well, the frame

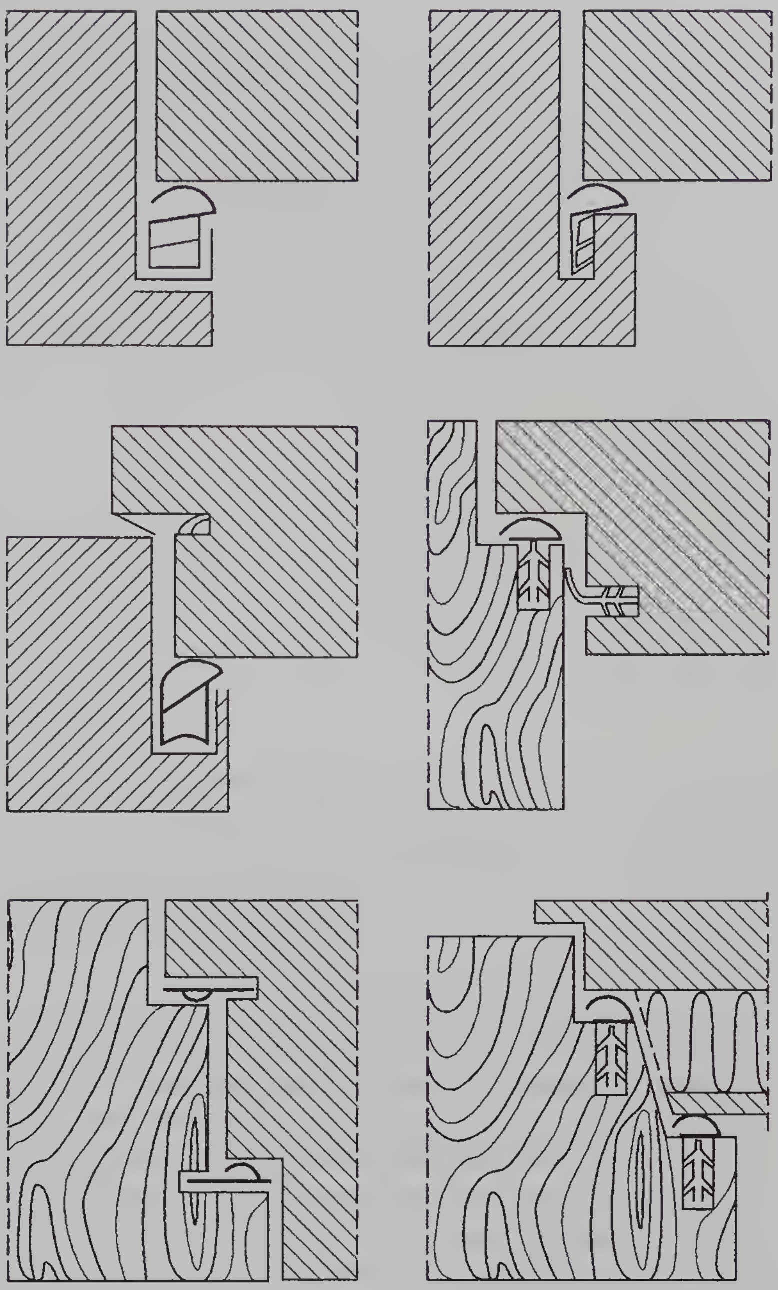


Fig. 17 Examples of single- and double-sealed door openings.

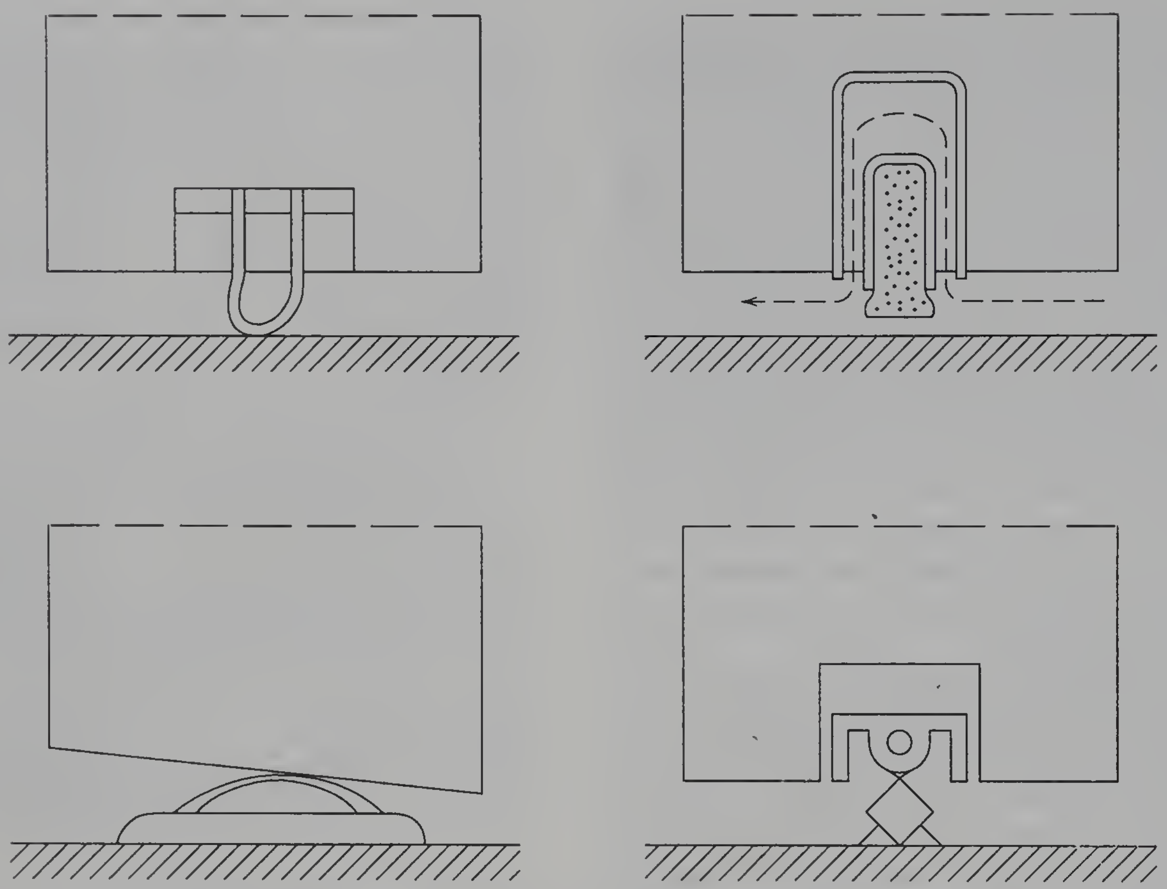


Fig. 18 Sealing of door openings on the ground without threshold.

must be sufficiently stiff. Otherwise, there is no significant dependence of sound insulation on the frame material whether of timber, metal, or plastic. Poor sealing increases sound transmission, especially in the high-frequency range. The maximum energy of traffic noise occurs at middle and low frequencies; therefore, for traffic noise, cavity depth and thickness of panes is more important than airtight joints.

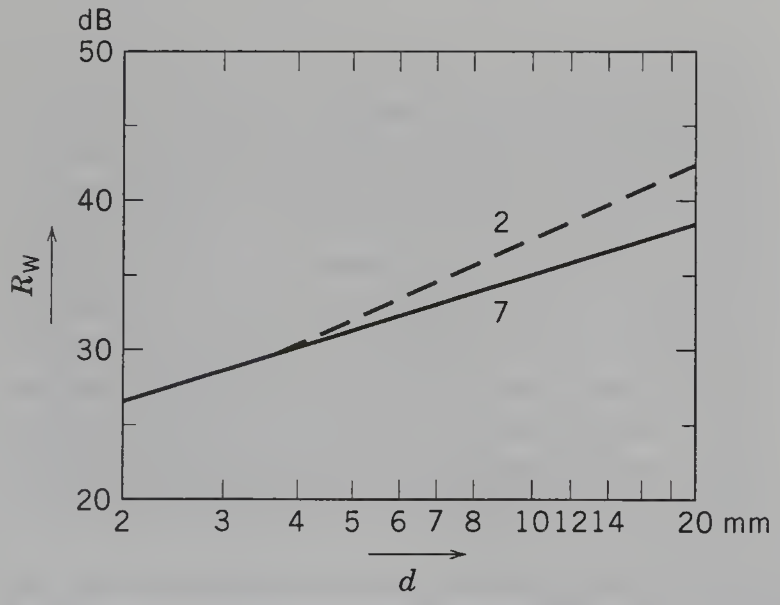


Fig. 19 Weighted sound reduction index R_w of single glazings vs. thickness d: (1) normal pane; (2) laminated pane.

Thermal glazing, consisting of two panes joined at the perimeter and placed into a single frame usually has a small cavity. Therefore, it would require very thick panes to obtain high sound insulation. A more practical solu-

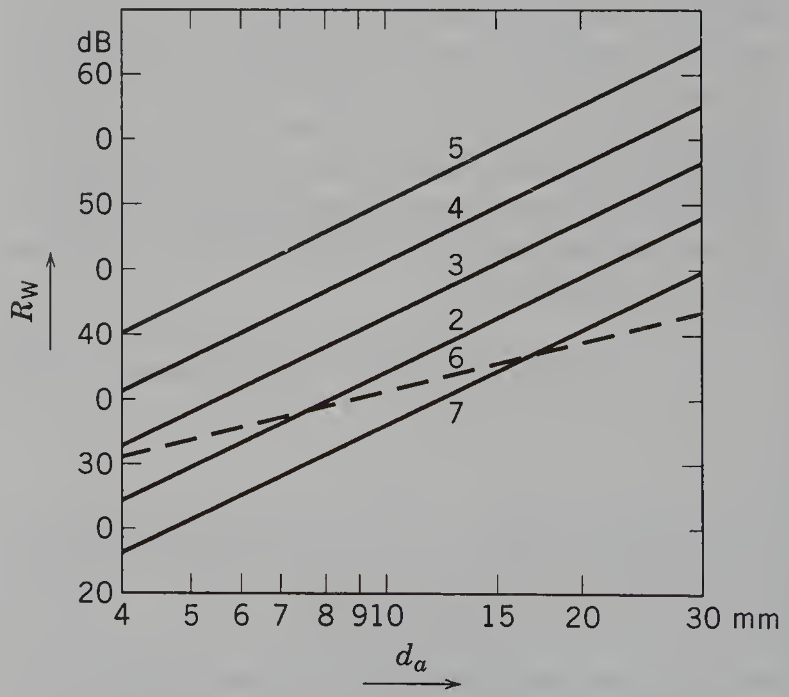


Fig. 20 Weighted sound reduction index R_w of double glazings vs. total thickness d_a of both panes. Distance: (1) 10 mm; (2) 20 mm; (3) 40 mm; (4) 80 mm; (5) 160 mm; (6) single glazing.

TABLE 8 Representative Values of Sound Insulation of Window Constructions

	Type of Window Construction					
	Single Frame with Single or Double Panes		Connected Double Frame with Single Pane and Double Pane or Two Double Panes		Large-Cavity Double Frame with Single Pane and Double Pane or Two Double Panes	
R'_w (dB)	Panes	Seals	Panes	Seals	Panes	Seals
<21	3	None	2 × 3	None	2 × 3	None
21–25	6 or 3/8/3	None	2 × 3	None	2 × 3	None
26–30	9 or 3/12/6	One	3/30/3	One	3/50/3	None
31–35	5/16/5 or 4/16/6	One	4/40/4 or 3/40/4	One	3/50/5 or 3/75/3	One
36–40	8/16/6 or 8/24/4	Two	8/50/4 or 8/50/6/12/4	Two	4/100/4 or 6/100/4/12/4	Two
41–45	No		10/60/8 or 8/60/8/12/4	Two	6/100/6 or 8/100/6/12/4	Two

tion is to add an additional pane (storm windows) to form a large cavity with the thermal glazing. An improvement of about 5 dB is possible by filling the cavity with a suitable heavy gas, such as sulfur hexafluoride (SF_6).

Table 8 gives some representative information about window constructions that meet different requirements on sound insulation. More information can be found in Refs. 48–50.

2.15 Flanking Transmission

Except in special laboratories for the measurement of sound transmission through building elements[11] there is practically always a certain amount of transmission of sound energy along paths other than the direct path shown in Fig. 1. These paths are called *flanking paths.*

Typically in buildings, in addition to the sound power W_2 transmitted through the separating wall, W_3, the sound power transmitted into the receiving room by flanking elements or by leaks, is significant. In this case, sound insulation is expressed by the *apparent sound reduction index*

$$R' = 10 \log \frac{W_1}{W_2 + W_3} \quad \text{dB}. \qquad (22)$$

The expression *apparent transmission loss* is also in use and is equivalent to the apparent sound reduction index.

Assuming that diffuse sound fields exist in the two adjacent rooms, the apparent sound reduction index may be evaluated from

$$R' = L_1 - L_2 + 10 \log \frac{S}{A} \quad \text{dB}, \qquad (23)$$

which is the same as Eq. (3) except that, because L_2 includes sound energy from all surfaces, we are dealing with the apparent sound reduction index. (In the case of staggered rooms, S is that part of the area of the partition common to both rooms.)

In typically furnished dwellings, the equivalent sound absorption area of sitting rooms and bedrooms is about 10–20 m^2; often the area of partition walls and floors has about the same magnitude. Thus the 10 log S/A term is very small, and there is little difference between apparent sound reduction index and level difference, $L_1 - L_2$.

The direct and various possible flanking transmission paths between two adjacent rooms are shown in Fig. 21. The sound power transmitted into the receiving room can be assumed to consist of the sum of the following components:

W_{Dd}, which has entered the partition directly and is radiated from it directly (= W_2 in Fig. 1);

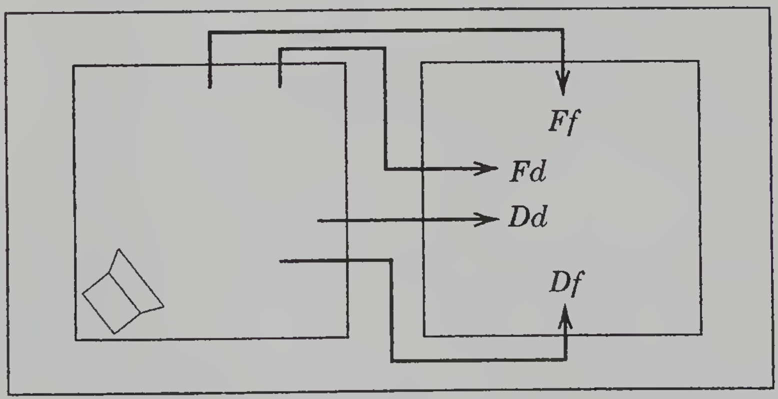


Fig. 21 Transmission paths between two adjacent rooms: *Dd* is the main path; *Ff*, *Df*, and *Fd* are flanking paths.

- W_{Df}, which has entered the partition directly but is radiated from flanking constructions;
- W_{Fd}, which has entered flanking constructions and is radiated from the partition; and
- W_{Ff}, which has entered flanking constructions and is radiated from flanking constructions.

The total power transmitted along flanking paths is thus

$$W_3 = W_{Df} + W_{Fd} + W_{Ff}. \qquad (24)$$

For each flanking path a sound reduction index (R_{Df}, R_{Fd}, R_{Ff}) can be defined, for example,

$$R_{Ff} = 10 \log \frac{W_1}{W_{Ff}} \quad \text{dB}. \qquad (25)$$

The index R_{Ff} (also denoted R_L) is of special interest because it depends only on flanking constructions. Improvements in the sound insulation of the primary separating element have little or no influence on R_{Ff}. It can be thought of as an upper limit to the sound insulation in any construction type.

The apparent sound reduction index R' between two adjacent rooms can be calculated from

$$R' = -10 \log \left(10^{-R/10} + \sum_i 10^{-R_i/10} \right) \quad \text{dB}, \qquad (26)$$

where the index i stands for Df, Fd, and Ff as appropriate.

At each junction in the building there is a certain amount of attenuation that depends on the type of junction, mass ratios, and damping in the various components. Some methods for calculating the effect of flanking have been proposed,[51,52] but the subject is still developing. Probably the most accurate results are obtained using Statistical Energy Analysis techniques. In a concrete apartment building about 50% of the incident sound power is transmitted to the adjacent rooms along flanking paths. Great attention has to be paid to flanking paths if high sound insulation is to be obtained in a building.

In the case of a lightweight wall perpendicular to a heavy floor, the attenuation of vibration in the floor due to the wall is very low. The importance of this is clearly evident in Fig. 22 for the example of a double-leafed lightweight wall of high sound insulation (more than R_w = 55 in the laboratory) resting on a concrete slab. Depending on the mean mass per unit area of the slab, the apparent sound reduction index of the wall increases from 42 up to 54 dB. In the vertical direction the trans-

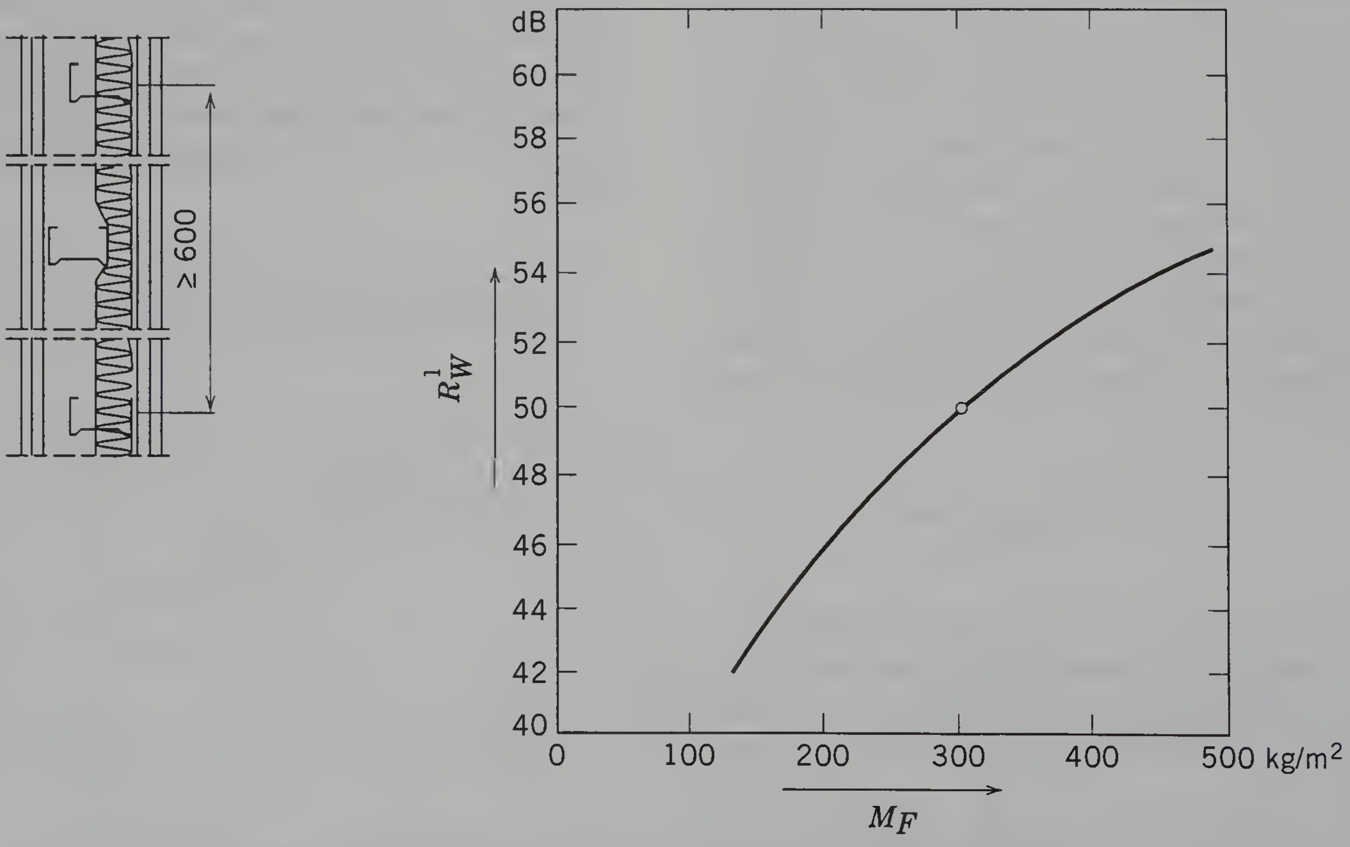


Fig. 22 Weighted apparent sound reduction index R'_w of a lightweight double wall vs. mean mass per unit area M_F of the flanking constructions.

mission is controlled by the floor and the lightweight walls have little effect. Lightweight wall systems are often used in apartment buildings. Thus structure-borne sound propagation in the floor slab and side walls is of great importance in such a building.

2.16 Multielement Partitions, Apertures, and Leaks

The total sound energy passing through each component of a multielement partition depends on the product of the area and sound transmission coefficient of each component. The equation governing the process is

$$A_t\tau_t = \sum_{i=1}^{N} A_i\tau_i, \qquad (27)$$

where A_i and τ_i represent the area and transmission coefficient of the ith component of the partition. The quantities A_t and τ_t represent the total area and effective composite transmission coefficient of the multielement partition. The transmission coefficients for the individual components may be calculated from their sound reduction indices by rearranging Eq. (2) as

$$\tau = 10^{-R/10}. \qquad (28)$$

The sound insulation of a multielement building partition is usually controlled by the component with the lowest sound insulation. In case of an opening, the sound reduction index is zero (the transmission coefficient is 1), and only the ratio of the area of the opening to the area of the rest of the partition determines the sound insulation. An opening of 10%, for example, an open door or open window in a wall, limits the total sound reduction index to 10 dB.

The approximation $\tau = 1$ is not correct, however, for small apertures and narrow slits with dimensions small compared with the wavelength. The sound transmission coefficient of those openings can be much higher than 1. Due to resonant effects, this is especially evident for depths of apertures and slits near the half of the wavelength. Apparent transmission coefficients $\tau > 100$ have been measured.[53,54]

Figure 23 shows an example of measurement results[55] of sound insulation for a plate with slits of various breadths. The decrease of sound insulation in the high-frequency range is typical. Sound transmission through an aperture depends on its position in the building element. The sound pressure at an edge of wall is higher than that in the middle, and the sound transmission through an aperture there is correspondingly higher.

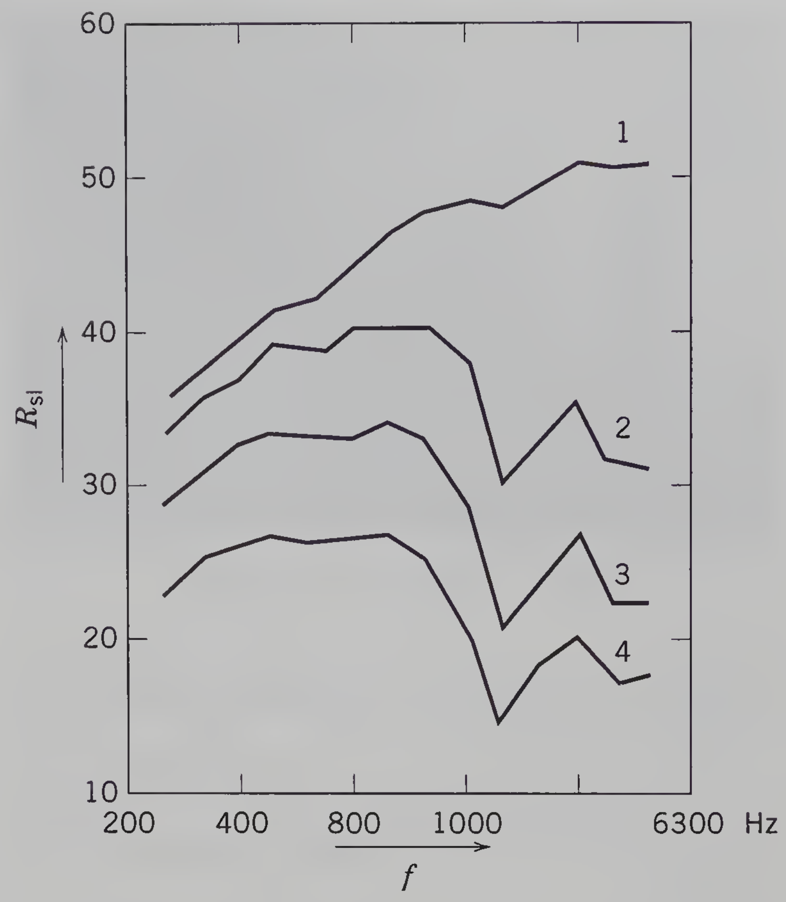


Fig. 23 Sound reduction index R_{sl} of slits (length 1 m) of various breadths relative to 1 m^2 area in a plate of 75 mm thickness: (1) without a slit; (2) breadth 2 mm; (3) breadth 3 mm; (4) breadth 7 mm.

3 IMPACT SOUND INSULATION

3.1 Definitions

Impact sound was defined above as a special kind of structure-borne sound. Perhaps the most common source is footsteps, but there are many others, such as furniture being moved and cleaning and other equipment operating directly on the surface of floors, stairs, and landings. Impact sound is radiated into rooms below, but besides this generally important case, horizontal transmission between apartments or from staircases or even transmission to the room above can cause annoyance.

The definition of impact sound pressure level is based on excitation using a standard tapping machine.[56–59] This machine, shown in Fig. 24, consists of five hammers placed along a line with a distance between adjacent hammers of 100 mm. The mass of each hammer is 500 g, its diameter is 30 mm, and it falls freely from a height of 40 mm. The time between successive impacts is 100 ms. For determination of impact sound insulation, the tapping machine is placed on the floor under test in different positions. In the receiving room the average sound pressure level L_i and the reverberation time are measured in

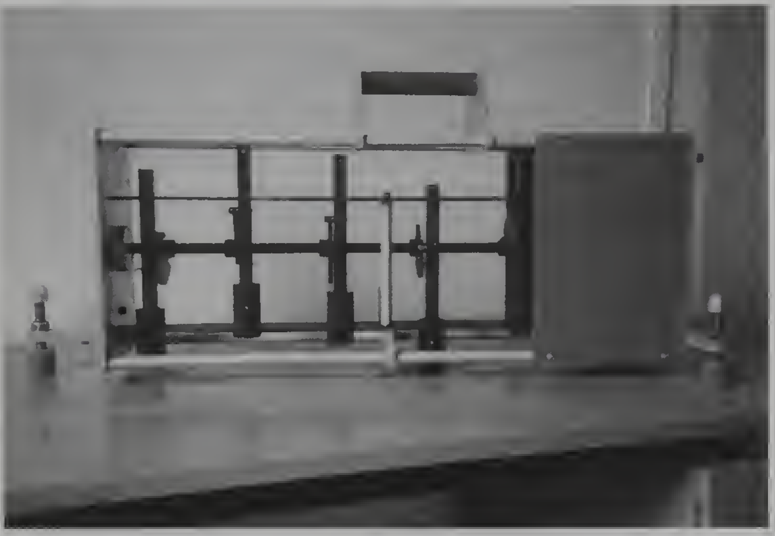


Fig. 24 Standard tapping machine.

one-third-octave bands, and the normalized impact sound pressure is calculated.

The *normalized impact sound pressure level* L_n is the level corresponding to a reference value of equivalent sound absorption area in the receiving room. It is, by definition,

$$L_n = L_i + 10 \log \frac{A}{A_0} \quad \text{dB}, \qquad (29)$$

where A is the equivalent sound absorption area in the receiving room and A_0 is the reference equivalent sound absorption area (10 m^2). Where it is uncertain that results are obtained without flanking transmission, the normalized impact sound pressure level is denoted by L'_n.

The impact sound insulation of concrete floors is generally unsatisfactory unless improved by a soft floor covering or a floating floor. Such measures cause a reduction of impact sound pressure level denoted by

$$\Delta L = L_{n0} - L_n \quad \text{dB}, \qquad (30)$$

where L_{n0} is the normalized impact sound pressure level in the absence of the floor covering and L_n is the normalized impact sound pressure level when the floor covering is in place. If a floating floor (see Section 3.5) is covered by a flexible covering, both ΔL values can be added in each one-third-octave band. That is not possible when two soft coverings are put one upon the other.

3.2 Single-Number Ratings

The impact sound pressure levels defined above depend on frequency, and standards require that they be measured in the frequency range 100–3150 Hz. To obtain a single-number rating of impact sound, a reference curve is used[60,61] in a procedure analogous to that for airborne sound insulation. The reference curve is shown in Fig. 25 compared with examples of measured curves for a wood joist floor with and without a concrete topping. In contrast to airborne sound, an unfavorable deviation at a particular frequency occurs when the measured value is higher than the reference value. The weighted values are impact insulation class, denoted by IIC,[61] and the weighted index $L_{n,w}$.[60] In this case, there is no difference between the ISO and the ASTM reference curves or fitting procedures. The ASTM procedure includes one more step, however. When the fitting is complete, the IIC is calculated by subtracting the value of the reference contour at 500 Hz from 110 (STC = 110 $-L_{n,w}$). Thus the IIC rating, similar to the STC and R_w, *increases* as the impact sound insulation improves. The $L_{n,w}$ rating *decreases* as the impact sound insulation improves.

In Fig. 25, the reference contour is fitted to the floor with the concrete topping. The single-number rating is controlled by the high-frequency levels. Despite the increased weight, the impact sound insulation is reduced. There is substantial improvement at low frequencies, however, and the addition of a soft covering on the concrete or a resilient support between it and the plywood would greatly improve the rating.

Since its introduction, there has been considerable debate about the usefulness of ISO tapping machine data

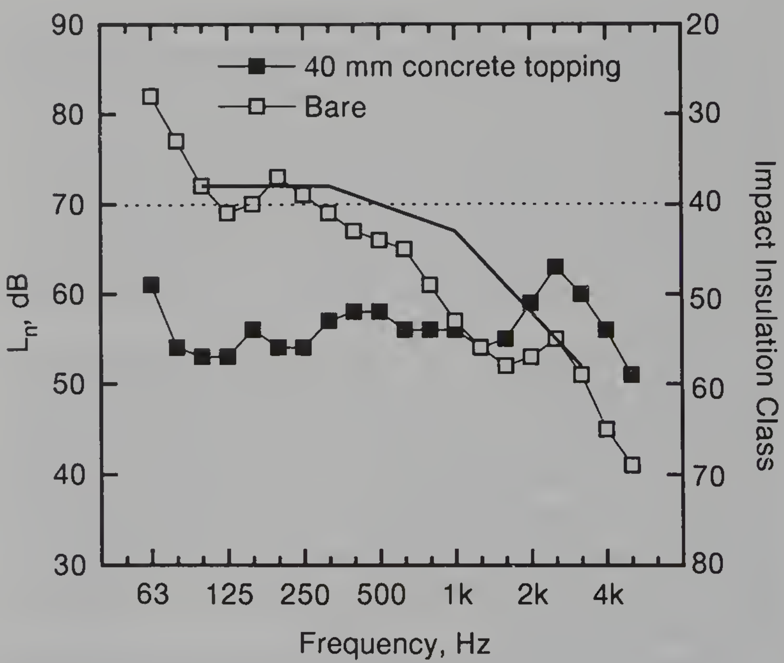


Fig. 25 Curve of reference values for impact sound compared with measurement results. The floor consists of 16-mm plywood resting on 240-mm deep wood joists (400 mm on center). The space between the joists contains a layer of glass fiber 90 mm thick. The ceiling is 16-mm gypsum board suspended on resilient metal channels (□). (■) For the same floor with a layer of concrete 40 mm thick placed on top. The reference contour is fitted to this curve.

obtained on different types of floors.[62–68] The debate continues.

In many European countries, most apartment buildings are constructed from concrete. There are several standard test procedures that make it more convenient to evaluate concrete floors and toppings for such construction. The procedures are not always in use in other countries but are of some interest.

The reduction of impact sound pressure level by a floor covering is, to a large extent, dependent on the floor on which it has been measured. Therefore, to get a single-number rating for the reduction of impact sound pressure due to a floor covering, the measured values of ΔL have to be related to a reference floor.[69] The reference floor is defined by the curve $L_{n,r,0}$, which can be found in Ref. 60. Procedures for making the measurements and calculations are dealt with in the appropriate standards.[56,60]

The resulting curve $L_{n,r} = L_{n,r,0} - \Delta L$ has to be evaluated following the standard procedure[60] to get the weighted normalized impact sound pressure level $L_{n,r,w}$. Thus the weighted impact sound improvement index ΔL_w of the floor covering can be calculated by

$$\Delta L_w = L_{n,r,0,w} - L_{n,r,w} = 78 - L_{n,r,w}, \qquad (31)$$

where $L_{n,r,0,w} = 78$ dB is the weighted normalized impact sound pressure level of the reference floor.

Another procedure is in use to evaluate the impact sound pressure level of bare concrete floors. These floors are invariably inadequate and any impact sound insulation is due to the covering. Therefore a more realistic rating of the bare concrete floor is obtained by taking into account the influence of a reference floor covering. The weighted sound improvement index of the reference floor[60] covering is $\Delta L_{r,w} = 19$ dB.

To evaluate a bare floor, the first step is to calculate the normalized impact sound pressure level of the floor with the reference covering from

$$L_{n,1} = L_{n,0} - \Delta L_r. \qquad (32)$$

From these values of $L_{n,1}$ a weighted value $L_{n,1,w}$ is calculated. The equivalent weighted normalized impact sound pressure level of the bare concrete floor can be found from

$$L_{n,w,\text{eq}} = L_{n,1,w} + \Delta L_{r,w} = L_{n,1,w} + 19 \text{ dB}. \qquad (33)$$

The definition of single-number ratings for floor coverings, ΔL_w, and bare concrete floors, $L_{n,w,\text{eq}}$, allows designers to calculate the weighted normalized impact pressure level of a completed floor, $L_{n,w}$, using the equation

$$L_{n,w} = L_{n,w,\text{eq}} - \Delta L_w. \qquad (34)$$

Thus requirements for impact sound insulation can be specified for floor coverings to match the bare concrete floor to be used.

The above deals with the most important situation of disturbance by impact sound where the receiving room is directly below the excited floor. For concrete constructions, when the receiving room is immediately adjacent on the same level, the impact sound insulation will be about 5 dB greater, about 10 dB greater if it is the second room on the same level, and about 10–20 dB greater, depending on flanking constructions, if the receiving room is above the source room. Such generalizations do not apply to lightweight joist construction using double-layer assemblies.

3.3 Concrete Floors

For a homogeneous reinforced concrete floor, results of calculations[17,70] and measurements[22] show a small increase of normalized impact sound pressure level with frequency (1.5 dB/octave). Doubling the thickness of such a floor causes an improvement of impact sound insulation of about 10 dB; nevertheless, in practice, concrete floors with no coverings or hard finishes adhered directly to the floor are never satisfactory.

This is also true for concrete floors with large or small cavities. Somewhat lower values of $L_{n,w,\text{eq}}$ of floors with small cavities (such as hollow-core concrete slabs), represented in Fig. 26, curve 2, are based on the increased thicknesses of these floors for the same mass per unit area as homogeneous floors. Higher values of $L_{n,w,\text{eq}}$ in curve 3 of Fig. 26 occur because of resonant effects in large hollows such as those in coffered floors with directly fastened ceilings and without absorbent material.

Suspended ceilings generally reduce impact and airborne sound transmission to a certain extent. About 5 dB reduction of $L_{n,w,\text{eq}}$ may be obtained, depending on the elasticity of the hangers. But in buildings, due to the flanking transmission in the walls, the improvement gained is generally not sufficient to warrant the cost.

3.4 Soft Floor Coverings

An obvious solution to the problem of impact sound is to reduce the excitation of the structural floor by covering it with a resilient layer. Of course, such layers do not significantly improve airborne sound insulation as well; any improvements are usually only at high frequencies.

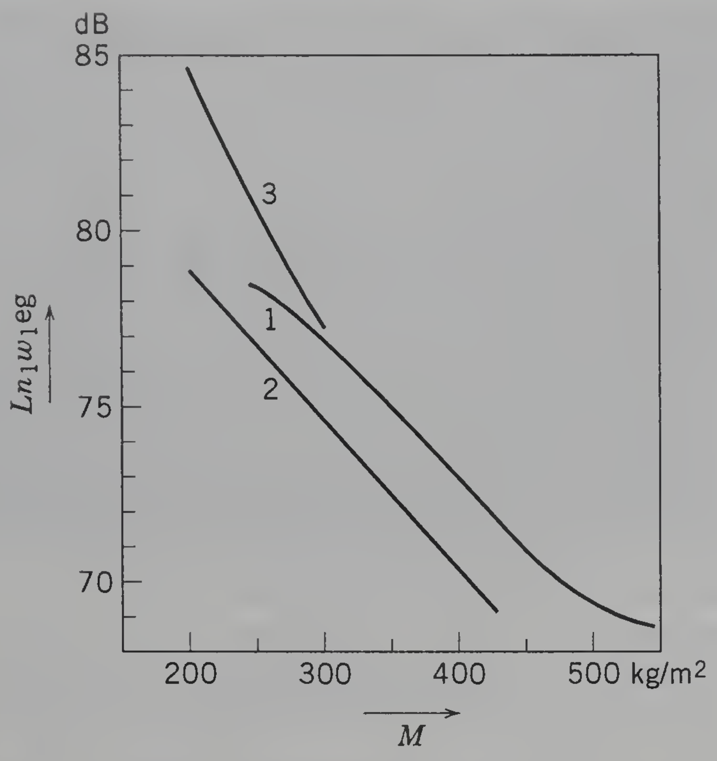


Fig. 26 Equivalent weighted normalized impact sound pressure level $L_{n,w,\mathrm{eq}}$ of concrete floors vs. mass per unit area M: (1) homogeneous material; (2) small cavities; (3) large hollows.

Soft floor coverings are effective in reducing the impact sound pressure level due to the standard tapping machine beginning at a resonant frequency f_0 given by

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{A_h E}{mh}}$$

$$= 5.98 \times 10^{-3}\sqrt{\frac{E}{h}} \quad \text{Hz}, \qquad (35)$$

where A_h is the striking area of the tapping machine hammer (in square metres), E is the dynamic Young's modulus of the covering (in newtons per metre), m is the mass of the hammer (in kilograms), and h is the thickness of the covering (in metres).

To estimate the improvement due to a soft floor covering, the acceleration of a hammer of the tapping machine striking upon the covering has to be measured.[71] The weighted impact sound improvement index is

$$\Delta L_w = 71.5 - 17.3 \log a \quad \text{dB}, \qquad (36)$$

where a is the maximum acceleration of the hammer (in metres per square second).

The test specimen should have dimensions of about 200 × 200 mm and must be placed on a solid, heavy, horizontal surface.

Flexible layers of a uniform material such as plastic or rubber provide only a low weighted impact sound improvement index of about 5–10 dB. A higher impact sound insulation (ΔL_w about 15–20 dB) is obtained by use of soft floors composed of a hard upper layer (rubber, plastic, or linoleum) and a resilient lower layer (soft felt board, fiber board, or foamed plastics). Usually those layers are glued to the floor. Thus the influence of the adhesive on impact noise has to be considered. Variation of ΔL_w based on the type of adhesive may be around ±3 dB. The impact sound improvement index for a carpet depends on thickness (about 20 dB for 4–5 mm and about 25 dB for 6–8 mm). The combination of a soft underpad and a carpet can give an increase of around 50 dB or more.

3.5 Floating Floors

When a soft floor covering may not be used, the most practical means of obtaining high impact sound insulation is to use a floating-floor construction. This type of floor has the additional advantage that it will also improve airborne sound insulation.

As shown in Fig. 27, a floating floor consists of a load-distributing slab resting on the structural floor but separated from it by a resilient layer (e.g., mineral wool, glass fiber, foamed plastics). Instead of a continuous resilient layer, a number of resilient pads may be used. The cavity between the floating slab and the structural floor is still best filled with soft sound-absorbing material. Most floating floors are made of concrete (about 30–60 mm thick), but they may also consist of gypsum, asphalt, or wooden floorboards. In the latter case the boards are nailed to battens that rest on resilient material. Special attention has to be paid to avoid solid connections (sound

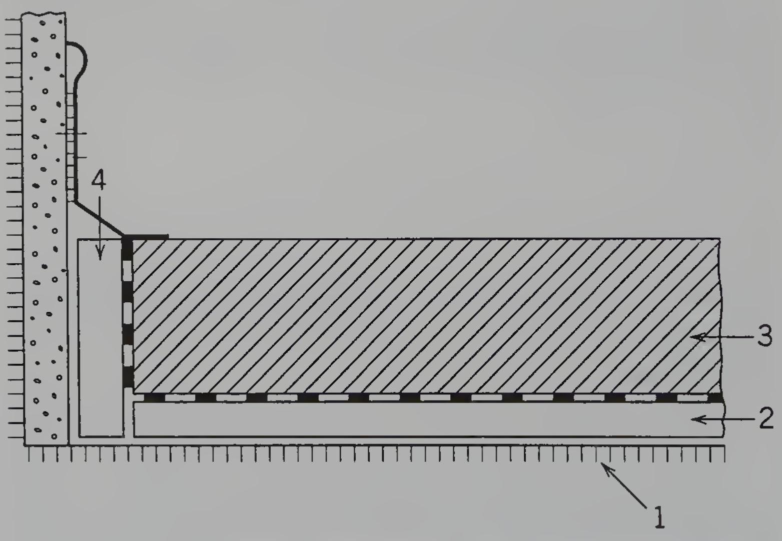


Fig. 27 Floating floor essentials: (1) structural floor; (2) resilient layer; (3) load-distributing slab; (4) insulation strip around edges.

bridges) between the floating slab and the structural floor. Therefore the elastic layer should be covered by plastic, roofing felt, building paper, or a similar material before pouring concrete. Resilient strips placed around all slab edges are necessary to avoid reduction of the impact sound insulation caused by sound transmission from the load-distributing slab to the walls.

As with other mass–spring–mass systems, the floating floor has a fundamental resonance frequency, f_{sl}, given by

$$f_{sl} = \frac{1}{2\pi}\sqrt{\frac{ns}{M_{sl}}}, \qquad (37)$$

where s is the dynamic stiffness of the elastic mounts (in newtons per metre), n is the number of mounts per unit area (for a continuous layer of material n is 1 and s is the dynamic stiffness per unit area of the material), and M_{sl} is the mass per unit area of the floating floor (in kilograms). As with other resonant systems, there is usually a decrease in sound insulation around the resonance frequency.

Where the floating slab is thick, rigid, and lightly damped, the tapping machine will generate an approximately reverberant bending wave field in the slab. In this case the improvement in impact insulation is

$$\Delta L_n \approx 10 \log \frac{2.3 M_{sl}\omega^3 \eta c_L h}{ns^2} \text{ dB}, \qquad (38)$$

where η is the loss factor in the slab, c_L is the longitudinal wave speed in the slab (in metres per second), and h is the thickness of the slab (in metres). This equation predicts an increase in ΔL of 30 dB/decade in frequency if the loss factor is independent of frequency. The dynamic stiffness per unit area of a resilient layer consists of the dynamic stiffness per unit area of the material and of that of the enclosed air, which includes the effects of airflow resistivity.[30,31,72] When separate resilient mounts are used, the stiffness of the entrapped air must also be considered.[73]

The improvements predicted in Eq. (38) are not always achieved in practice. Improvements in L_n for two types of floating slabs on two types of floors are shown in Fig. 28. When the slabs rest on the 150-mm-thick concrete, the negative ΔL_n around the resonance frequencies is evident. There are no negative values of ΔL_n when the slabs rest on the wood joist floor and, overall, the improvement from the floating slabs is not as great. The floating slab, F3, which is made of wood, has a more resilient surface than the concrete slab, so the improvements at high frequency are greater.

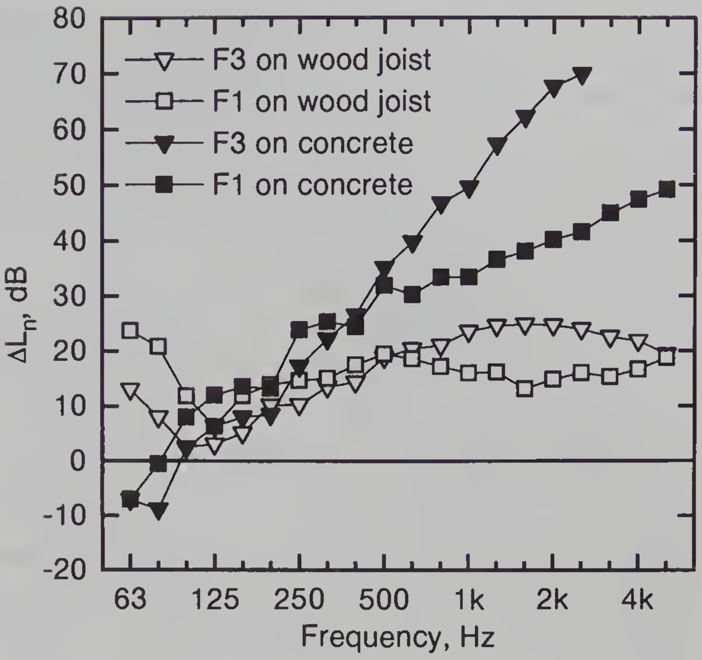


Fig. 28 Improvement in L_n' due to two types of floating slabs, F1 and F3, on 150-mm concrete slab and the wood joist floor described in Fig. 25. F1 is a 40-mm thick concrete slab floating on a glass fiber (25 mm thick, 50 kg/m^3). F3 is a 19-mm thick wood slab attached to 38-mm-thick strapping resting on glass fiber (25 mm thick, 114 kg/m^3).

The dependence of the weighted impact sound improvement index ΔL_w on the dynamic stiffness per unit area of the supporting resilient material is given in Fig. 29. It shows, for example, that a concrete slab with a mass of 50 kg/m^2 (about 25 mm thick) resting on a resilient layer with a dynamic stiffness per unit area of

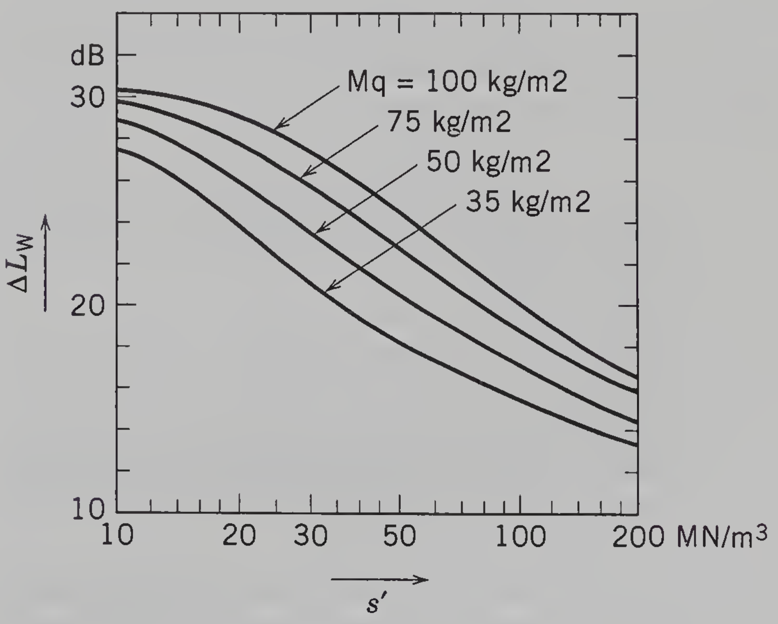


Fig. 29 Weighted impact sound improvement index ΔL_w of floating floors vs. dynamic stiffness per unit area s' of the resilient layer for different values of mass per unit area M_l of the load-distributing slab. The structural slab is assumed to be heavy and rigid.

30 MN/m · m^2 provides an improvement in the weighted impact sound index of 23 dB. From Fig. 26 we see that a homogeneous concrete floor 150 mm thick (330 kg/m^2) has $L_{n,w,\mathrm{eq}} = 76$ dB. Thus $L_{n,w} = 53$ dB (IIC = 57). A dynamic stiffness of 30 MN/m · m^2 can be obtained by using a resilient layer of mineral wool or glass fiber about 10 mm thick (density 100–150 kg/m^3).

The dynamic stiffness per unit area of elastic layers to be used under floating floors can be measured by a resonance method.[37] In this method, the fundamental resonant frequency of the vertical vibration of a spring and mass system is measured. The spring is the test specimen of the resilient material under test (200 × 200 mm) and the mass is a load plate placed on it. The load plate is excited by sinusoidal signals. Once the dynamic stiffness per unit area of the elastic layer has been determined, the weighted impact sound improvement index of the floating floor can be found in Fig. 29.

3.6 Joist and Truss Floors

The same factors that control sound transmission through double-layer walls are as important in double-layer floors. To avoid having the floor and ceiling layers rigidly connected usually requires separate joists or a resilient suspension system for the ceiling.

Double-layer joist or truss floors are generally much lighter than solid concrete floors, and it is very difficult to reduce the transmission of impact sound at the low frequencies below 100 Hz. The IIC or $L_{n,w}$ rating might meet all code requirements but building occupants may still complain about thumping sounds from people walking.

As much weight as can be tolerated and some kind of floating floor offer the best available solution. Floating floors take three main forms:

1. A solid slab of concrete (typically 40–50 mm thick), wood, or some other material resting on a resilient sheet.
2. Wooden flooring nailed to battens to form a raft. The raft rests on a sheet of resilient material laid on a subfloor with the battens directly above and parallel to the floor joists. Instead of a sheet of resilient material, pads or strips might be placed under the battens and sound-absorbing material placed in the cavity between the battens.
3. Where allowed by building codes, a wood raft may also rest on resilient layers draped over the joists with no intervening subfloor.

In all cases the raft must not be nailed down to the joists at any point or contact the side walls anywhere. It should be isolated from the surrounding walls to avoid flanking transmission, for example by leaving a gap round the edges that can be filled with resilient material and covered by the skirting (baseboard).

A heavy pugging (a loose filling usually poured between the joists) is advantageous for adding mass to a floor. In older wooden joist constructions, puggings are placed on separate pugging boards inserted between the joists. Such puggings may consist of granular or fibrous materials such as sand or ashes. In reconstructed wooden joist floors often these separate pugging boards have been removed, in which case the floor should be reconstructed to have a resiliently supported ceiling with sound-absorbing material placed in the cavity.

The impact sound insulation of wooden joist floors cannot be calculated by dividing into the acoustical effects of floor coverings and structural floors, as is the case with concrete. The weight and structure of the upper layer can vary quite widely, and it is best to find measured data for specific types of floors. Figure 30 shows airborne and impact sound insulation for some types of wooden joist floors that might be used in Europe. Table 9 shows data for construction more typical of North America.

Table 9 makes it plain that it is relatively simple to get high airborne sound insulation ratings with joist construction. Similar results would be expected for wood I-beams or trusses. The joists in the table were 240 mm deep. Deeper joists, trusses, or I-beams would give better results. High impact sound ratings are not so easy to attain. The impact sound ratings are very much controlled by the resilience of the surface layer or by the presence of a soft layer under a floating slab. An impact insulation class greater than 55 with a hard-surfaced floor results in an STC rating over 60. The last floor in the table serves as a warning. It shows the price to be paid if the floor layer is rigidly connected to the ceiling and the danger of allowing small air cavities in the ceiling. Resilient metal channels should never be used between layers of gypsum board in a floor or a wall.

3.7 Standard Tapping Machine Compared to a Walker

The standard tapping machine test procedure does not deal with impact sound generated at frequencies below 100 Hz. With lightweight joist floors, low-frequency impact sound can be very annoying. Figure 31 shows tapping machine and walker sound levels generated under a 150-mm concrete slab with and without a carpet. For the tapping machine, for all but the very low frequencies, the carpet has a very significant effect. The walker was wearing shoes with rubber-tipped heels and so generated less high-frequency sound than the tapping

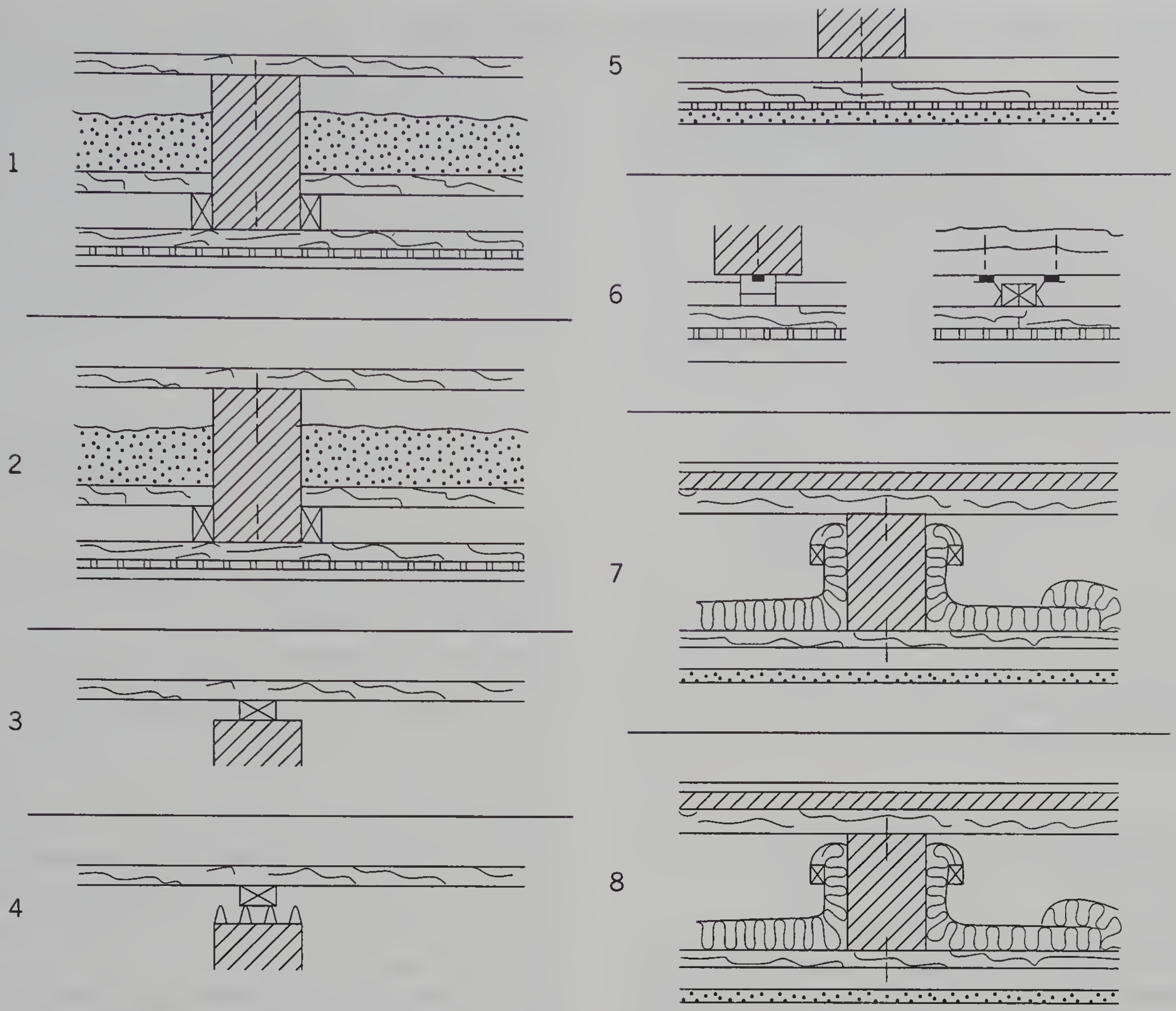


Fig. 30 Airborne and impact sound insulation of wooden joist floors. (1) Wooden floorboards (25 mm) directly nailed to the joists (200 mm); pugging of slag on separate pugging boards; plastering on expanded metal below ceiling boards, directly fastened to the joists: $R'_w = 45$ dB; $L'_{n,w} = 68$ dB. (2) Like 1, but mineral wool (60 mm) instead of pugging; gypsum board (12.5 mm) instead of plastering: $R'_w = 49$ dB; $L'_{n,w} = 65$ dB. (3) Like 1, but floorboards nailed to battens; $R'_w = 49$ dB; $L'_{n,w} = 63$ dB. (4) Like 3, but raft resting on layers of mineral wool (10 mm): $R'_w = 51$ dB; $L'_{n,w} = 58$ dB. (5) Like 1, but ceiling fixed to battens perpendicular to the joists: $R'_w = 49$ dB; $L'_{n,w} = 63$ dB. (6) Like 3, but ceiling fixed with resilient metal hangers: $R'_w = 51$ dB; $L'_{n,w} = 58$ dB. (7) Wooden floorboards (25 mm) floating on mineral wool (20 mm); wooden floorboards (25 mm) directly nailed to the joists (200 mm); absorbent material (50 mm); ceiling of gypsum board (2 × 12.5 mm), fixed with crossing battens: $R'_w = 53$ dB; $L'_{n,w} = 53$ dB. (8) Like 7, but floating concrete slab (40 mm) instead of wooden floorboards: $R'_w = 55$ dB; $L'_{n,w} = 51$ dB.

machine on the bare floor, but the addition of the carpet still makes a significant improvement for frequencies over 100 Hz. At the very low frequencies, the carpet provides no improvement, but the sound levels there are relatively low anyway.

In contrast, the same kind of data are shown in Fig. 32 for a wood joist floor. Again the carpet makes little difference at low frequencies, but for this floor the low-frequency levels are much higher and more likely to be annoying. The data for the walker also show indications of floor creaks that are not present in the more rigid concrete floor or with the tapping machine.

3.8 Staircases

In apartments, there is usually no need for special measures to control impact sound insulation in stairways or landings if staircases will not be directly adjacent to sitting rooms or bedrooms. There will be at least two junctions between the stairs and sensitive areas, so impact

TABLE 9 Impact and Airborne Sound Ratings for Selected Floors

Rating Code	Floor Surface	Ceiling Surface	IIC	STC	R_w
A	One layer of 15-mm oriented strand board (OSB)	One layer of 16-mm gypsum board	46	51	51
B	Two layers of 15 mm OSB	As A	47	55	56
C	One layer of 15 mm OSB	Two layers of 16-mm gypsum board	49	55	55
D	Two layers of 15 mm OSB	As C	53	60	60
E	One layer of 16 mm plywood	As C	41	52	53
F	As E + 13 mm hardwood on 3-mm shredded rubber	As C	50	57	61
H	As E + 9 mm wood parquet, 40-mm normal concrete and 1-mm-thick asphalt paper	As C	55	69	69
I	As E + 9 mm wood parquet, 40-mm cellular concrete and 1-mm-thick asphalt paper	As C	53	67	67
J	One layer of 16-mm pylwood or OSB	Two layers of 16-mm gypsum board with 13-mm resilient metal channels between	31	38	38

Note: Constructed using 38 × 240 mm wood joists, 400 mm on center (oc). All gypsum board layers are supported on resilient metal channels spaced 600 mm oc.

sound insulation in the horizontal direction should be sufficient (see Section 2.15) even for stairs with a concrete surface.

Otherwise, gaps between the stairway and the landing and staircase wall are a suitable means of improving impact sound insulation. Gaps may be open or filled with any resilient material and, naturally, can be used only where structural considerations for the stairway or landing do not necessitate support on the partition wall. Gaps should be covered to prevent accumulation of debris that might bridge the gap.

In some cases, the only option available is an elastic support of the stairways or landings. Rubber pads (thickness not less than 20 mm) may be a suitable material for

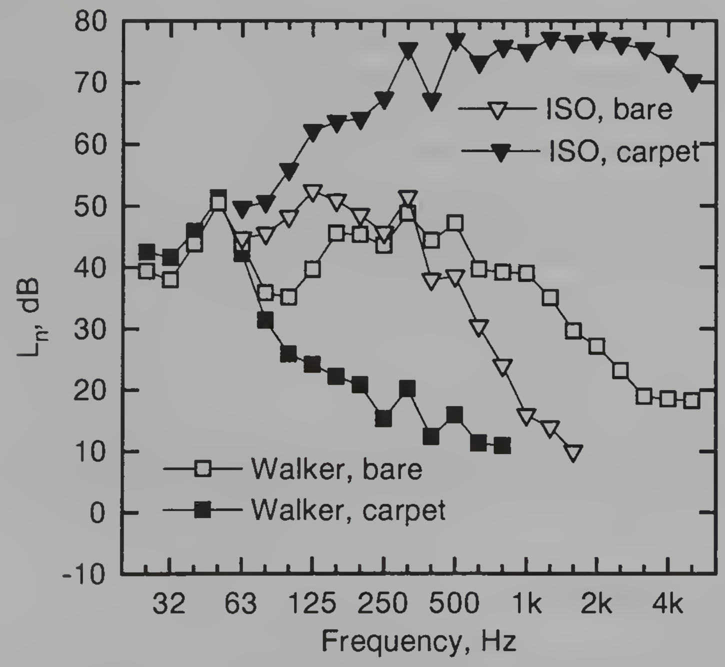


Fig. 31 Impact sound transmission through a 150-mm-thick concrete floor with and without a carpet and underpad. The levels are generated by a person and ISO standard tapping machine.

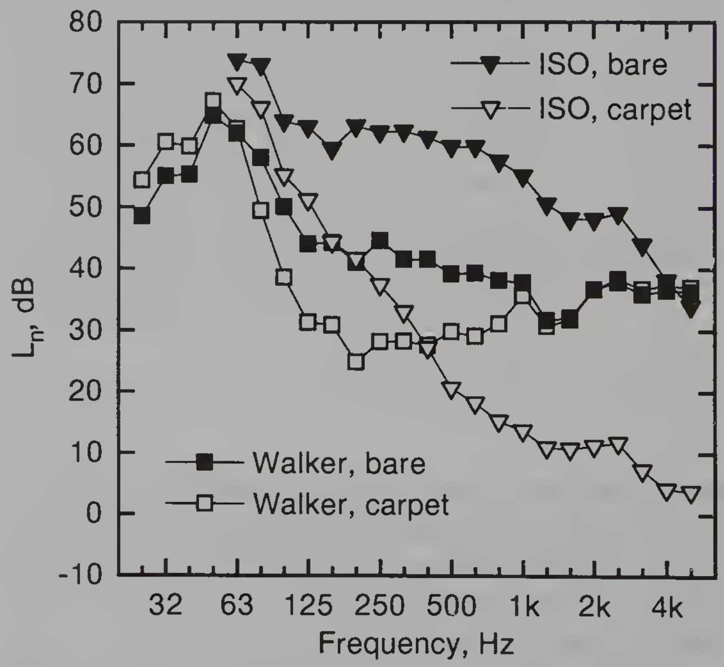


Fig. 32 Impact sound transmission through a wood joist floor with and without a carpet and underpad. The floor is similar to that in Fig. 25 but with two layers of 16-mm gypsum board and 270 mm of glass fiber batts in the cavity. The levels are generated by a person and the standard tapping machine.

this purpose. A soft or floating floor can also be used for stairs and landings. Because of the attenuation across junctions in the horizontal direction, any requirements on weighted impact sound improvement index may be diminished by about 5 dB. Nevertheless, construction of an effective floating floor needs the greatest care for it to work properly. In practice it is best to combine a floating floor on the landings with an elastic support of stairways upon the landings and gaps between stairways and walls of the staircase.

4 ACOUSTICAL REQUIREMENTS IN BUILDING CODES

National building codes and standards usually contain requirements for sound insulation aimed at meeting the demands of different functions of rooms and buildings. Internationally, the minimum requirements may be of similar magnitude, but in detail there are important differences. In apartment houses, especially, a compromise is necessary between the demand for a sufficiently quiet atmosphere on the one hand and economic practicalities on the other, the latter causing national differences. Therefore this section only deals with the kind of requirements and gives some examples.

Of major importance are the requirements on interior building elements that separate two apartments, but there are also requirements between particular rooms in public and office buildings. For this, the apparent weighted sound reduction index (R'_w), or STC, and the weighted normalized impact sound pressure level ($L'_{n,w}$), or IIC, are in use (see Sections 2.6 and 3.2).

For example, in Germany the requirements are $R'_w \geq$ 53 dB for partition walls and $R'_w \geq 54$ dB and $L'_{n,w} \leq 53$ dB for floors between dwellings.[74] These values, though corresponding with a good and fairly costly sound insulation, do not ensure that loud voices or music, household machinery, or children jumping in the adjacent flat cannot be heard. Therefore two levels giving a higher level of privacy are recommended[75]: $R'_w \geq 56$ dB and $R'_w \geq 59$ dB for partition walls and $R'_w \geq 57$ dB, $L'_{n,w} \leq 46$ dB and $R'_w \geq 60$ dB, $L'_{n,w} \leq 39$ dB for floors.

To protect against noise from staircases in apartment houses, requirements on impact sound insulation of stairs and landings ($L'_{n,w} \leq 58$ dB) and on airborne sound insulation of doors ($R'_w \geq 37$ dB, without corridors) have to be met.[59]

Higher demands have to be met if dwellings are adjacent to noisy rooms (restaurants, shops, workshops, heating stations, etc.). The required sound insulation depends on the average equivalent sound pressure level in the source room. For example, for $L_{\text{eq}} \leq 80$ dB(A), $R'_w \geq$ 57 dB is reqired; for $L_{\text{eq}} \leq 85$ dB(A), $R'_w \geq 62$ dB is required; and for $L_{\text{eq}} \leq 95$ dB(A), $R'_w \geq 72$ dB is required.[74]

Requirements for exterior walls in apartment houses and public buildings, especially for windows, depend on outside noise level and spectrum and on the permissible inside noise level [e.g., $L_{\text{eq}} \leq 30$ dB(A) in bedrooms and in living rooms]. The problem then is to select an appropriate value of R'_w for the facade when given an equivalent sound pressure level $L_{\text{eq},o}$ caused by outside noise (about 1–2 m in front of the building) and a legally permissible equivalent sound pressure level $L_{\text{eq},i}$ in the receiving room (both A weighted). The equation used is

$$R'_w = L_{\text{eq},o} - L_{\text{eq},i} + 10 \log \frac{S}{A} + C \quad \text{dB}, \qquad (39)$$

where S is the area of the facade, A is the equivalent sound absorption area of the receiving room (approximation for living rooms $A \approx 0.8 \times$ floor area), and C is a coefficient taking into account the spectrum of the outside noise (in decibels). The following values of C can be used in practice[75]:

Railway noise	0–3 dB
Road traffic noise	3–6 dB
Aircraft noise	6–8 dB

For typical housing, outside equivalent sound pressure levels $L_{\text{eq},o}$ of 61 to 65 dB(A) from road traffic noise require an apparent weighted sound reduction index of $R'_w = 40$ dB for the facade. This value increases or decreases in steps of 5 dB corresponding to changes of 5 dB(A) of the outside level.

For most housing construction the sound insulation of the facade generally depends entirely on the windows. Sound transmission through the wall has to be considered only in few cases, for example, when an exterior wall is a lightweight sandwich component or the like, which will be found mainly in office buildings. The required sound insulation of windows can be calculated, taking into account the area ratio of the window to the whole facade (see Section 2.16).

Demands on the airborne sound insulation in industrial buildings are also influenced by kind of noise source and by permissible sound pressure level. Usually requirements are given in A-weighted sound pressure level. For transmission to interior building elements the necessary apparent weighted sound reduction index can be found using Eq. (23) (see Section 2.15).

If the concern is sound transmission to a receiving point outside an industrial hall, the following equation can be used to get the necessary sound reduction index of the facade:

$$R' = L_i - L_p + 10\log \frac{S}{r^2} - 14 \text{ dB}, \qquad (40)$$

where L_i is the sound pressure level within a frequency band (octave or one-third octave) inside the building, L_p is the permissible sound pressure level in a frequency band at the receiving point, S is the area of the radiating facade, and r is the distance of the receiving point to the facade ($r \geq \sqrt{S}$). For facades consisting of components of different sound insulation, the information in Section 2.16 has to be used.

4.1 Sources of Information on Sound Transmission

Many manufacturers provide collections of sound transmission data using their products. As well, other agencies have printed collections of information.$^{76-80}$ When comparing nominally identical structures, in such publications it is important to be sure that all details are identical. Small changes in mass, screw, or stud spacing or some other detail can have a large effect on the sound transmission. The density of gypsum board, for example, is not a constant.

REFERENCES

1. Lord Rayleigh, *The Theory of Sound*, Dover, New York, 1945.
2. L. Cremer, "Theorie der Schalldammung dunner Wande bei schrägem Einfall," *Akustische Zeitschrift*, Vol. 3, 1942, pp. 81–93.
3. A. London, "Transmission of Reverberant Sound through Double Walls," *J. Acoust. Soc. Am.*, Vol. 44, 1950, pp. 270–278.
4. H. Cremer and L. Cremer, "Theorie der Entstehung des Klopfschalls," *Frequenz*, Vol. 2, 1948, p. 61.
5. A. London, "Transmission of Reverberant Sound through Single Walls," *J. Res. Natl. Bureau Stand.*, RP1998, Vol. 42, June 1949, p. 605.
6. A. London, "Transmission of Reverberant Sound through Double Walls," *J. Res. Natl. Bureau Stand.*, RP2058, Vol. 44, January 1950, p. 605.
7. W. Fasold, E. Sonntag, and H. Winkler, *Bauphysikalische Entwurfslehre, Bau- und Raumakustik*, Verlag für Bauwesen, Berlin, 1987.
8. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Laboratory Measurements of Airborne Sound Insulation of Building Elements," ISO 140-3, 1978.
9. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Measurement of Sound Insulation of Small Building Elements," ISO 140-10, 1989.
10. American Society for Testing and Materials, "Standard Test Method for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions," ASTM E90, 1990.
11. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Requirements for Laboratories," ISO 140-1, 1990.
12. "Bauakustische Prüfungen, Luft- und Trittschalldamm-Messungen an Bauteilen," Deutsches Institut für Normung, DIN 52210, T.2, 1984.
13. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Field Measurements of Airborne Sound Insulation between Rooms," ISO 140-4, 1978.
14. American Society for Testing and Materials, "Standard Test Method for Measurement of Airborne Sound Insulation in Buildings," ASTM E336, 1990.
15. M. J. Crocker, P. K. Raju, and B. Forssen, "Measurement of Transmission Loss of Panels by Direct Determination of Acoustic Intensity," *Noise Control Eng.*, Vol. 17, 1981, p. 6.
16. "Building Elements: Sound Insulation, Intensity Scanning under Laboratory Conditions," Nordtest method NT ACOU 084.
17. R. E. Halliwell and A. C. C., Warnock, "Sound Transmission Loss, Comparison of Conventional Techniques with Sound Intensity Techniques," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 2094–2103.
18. J. C. S. Lai, M. A. Burgess, P. P. Narang, and K. Miki, "Transmission Loss Measurements: Comparisons between Sound Intensity and Conventional Methods," *Proceedings of InterNoise 91*, 1991, p. 1029.
19. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Field Measurement of Airborne Sound Insulation of Facade Elements and Facades," ISO 140-5, 1978.
20. International Standardization Organization "Acoustics, Rating of Sound Insulation in Buildings and of Building Elements, Airborne Sound Insulation in Buildings and of Interior Building Elements," ISO 717-1, 1982.
21. International Standardization Organization "Acoustics, Rating of Sound Insulation in Buildings and of Building Elements, Airborne Sound Insulation of Facade Elements and Facades," ISO 717-3, 1982.
22. American Society for Testing and Materials, "Classification for Rating Sound Insulation," ASTM E413, 1994.
23. W. Fasold, Zur Bewertung der Schalldämmung von Fenstern, 5th International Congress of Acoustics, Lüttich, 1965.
24. American Society for Testing and Materials, "Classification for Determination of Outdoor–Indoor Transmission Class," ASTM E1332, 1994.
25. L. L. Beranek and I. L. Vér, *Noise and Vibration Control Engineering*, Wiley, New York, 1992.

26. D. A. Bies and C. H. Hansen, *Engineering Noise Control*, Unwin Hyman, London, 1988.
27. M. Heckl, "Investigations of Orthotropic Plates," *Acustica*, Vol. 10, 1960, p. 109 (in German).
28. L. L. Beranek and I. L. Vér, *Noise and Vibration Control Engineering*, Wiley, New York, 1992.
29. K. Gösele, "Verringerung der Luftschalldammung von Wänden durch Dickenresonanzen," *Bauphysik*, Vol. 12, 1990, pp. 187–191.
30. International Standardization Organization, "Acoustics, Material for Acoustical Applications, Determination of Airflow Resistance," ISO 9053, 1990.
31. American Society for Testing and Materials, "Standard Test Method for Airflow Resistance of Acoustical Materials. American ASTM Method C522, 1993.
32. B. H. Sharp, "A Study of Techniques to Increase the Sound Insulation of Building Elements," NTIS PB 222 829/4, U.S. Department of Housing and Urban Development, Washington, DC; "Prediction Methods for the Sound Transmission of Building Elements," *Noise Control Eng.*, Vol. 11, 1978, p. 53.
33. K. Gösele, "Prediction of the Sound Transmission Loss of Double Partitions (without Structureborne Connections)," *Acustica*, Vol. 45, 1980, p. 218 (in German).
34. G. Qiangguo and W. Jiquing, "Effect of Resilient Connection on Sound Transmission Loss of Metal Stud Double Panel Partitions," *Chinese J. Acoust.*, Vol. 2, No. 2, 1983, p. 113.
35. J. D. Quirt and A. C. C. Warnock, "Influence of Sound-Absorbing Material, Stud Type and Spacing, and Screw Spacing on Sound Transmission through a Double-Panel Wall System", *Proceedings Inter-Noise 93*, Vol. 2, 1993, pp. 971–974.
36. W. Loney, "Effect of Cavity Absorption on the Sound Transmission Loss of Steel Stud Gypsum Wallboard Partitions," *J. Acoust. Soc. Am.*, Vol. 49, 1971, p. 385.
37. P. P. Narang, "Effect of Fiberglass Density and Flow Resistance on Sound Transmission Loss of Cavity Plasterboard Walls," *Noise Control Eng. J.*, Vol. 40, No. 3, 1993, pp. 215–220.
38. K. Gösele and U. Gösele, "Einflußder Hohlraumdämpf ung auf die Steifigkeit von Luftschichten bei Doppelwänden," *Acustica*, Vol. 38, 1977, pp. 159–166.
39. "Summary Report for the Consortium on Gypsum Board Walls: Sound Transmission Loss Results," Internal Report IR693, Institute for Research in Construction, NRCC, Canada.
40. A. C. C. Warnock and J. D. Quirt, "Sound Transmission through Gypsum Board Walls," *Proceedings Inter-Noise 95*, 1995, pp. 727–730.
41. A. C. C. Warnock, "Sound Transmission through Concrete Blocks with Attached Drywall," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 1454–1463.
42. A. C. C. Warnock, "Sound Transmission Loss Measurements through 190 mm and 140 mm Blocks with Added Gypsum Board and through Cavity Block Walls," NRCC Internal Report IR586, 1990.
43. A. C. C. Warnock, "Sound Transmission through Two Kinds of Porous Concrete Blocks with Attached Drywall," *J. Acoust. Soc. Am.*, Vol. 92, 1992, pp. 1452–1460.
44. J. Enger and T. E. Vigran, "Transmission Loss of Double Partitions Containing Resonant Absorbers," *Proc. Inst. Acoust.*, Vol. 7, Part 2, 1985, p. 125.
45. J. M. Mason and F. J. Fahy, "The Use of Acoustically Tuned Resonators to Improve the Sound Transmission Loss of Double-Panel Partitions," *J. Sound Vib.*, Vol. 124, No. 2, 1988, p. 367.
46. A. C. C. Warnock, "Sound Transmission through Slotted Concrete Blocks with Attached Gypsum Board," *J. Acoust. Soc. Am.*, Vol. 94, No. 5, 1993, pp. 2713–2720.
47. J. D. Quirt, "Sound Transmission through Double Doors," *Can. Acoust.*, Vol. 14, No. 4, 1986, p. 3.
48. J. D. Quirt, "Measurements of the Sound Transmission Loss of Windows," Building Research Note No. 172, Institute for Research in Construction, NRCC, 1981.
49. J. D. Quirt, "Sound Transmission through Windows. I. Single and Double Glazing," *J. Acoust. Soc. Am.*, Vol. 72, No. 3, 1982, pp. 834–844.
50. J. D. Quirt, "Sound Transmission through Windows. II. Double and Triple Glazing," *J. Acoust. Soc. Am.*, Vol. 74, No. 2, 1983, pp. 534–542.
51. E. Gerretsen, "Calculation of Sound Transmission between Dwellings by Partitions and Flanking Structures," *Appl. Acoust.*, Vol. 12, 1979, p. 413.
52. E. Gerretsen, "Calculation of Airborne and Impact Sound Insulation between Dwellings," *Appl. Acoust.*, Vol. 19, 1986, p. 245.
53. M. C. Gomperts and T. Kihlmann, "The Sound Transmission Loss of Circular and Slit-Shaped Apertures in Walls," *Acustica*, Vol. 18, 1967, pp. 144–150.
54. G. P. S. Wilson and W. W. Soroka, "Approximation to the Diffraction of Sound by a Circular Aperture in a Rigid Wall of Finite Thickness," *J. Acoust. Soc. Am.*, Vol. 37, 1965, pp. 286–297.
55. K. Gösele and W. Schüle, Schall, Wärme, Feuchtigkeit 4. Auflage, Bauverlag Wiesbaden, Berlin 1977.
56. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Laboratory Measurements of Impact Sound Insulation of Floors," ISO 140-6, 1978.
57. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Field Measurement of Impact Sound Insulation of Floors," ISO 140-7, 1978.
58. American Society for Testing and Materials, "Standard Method of Laboratory Measurement of Impact Sound Transmission through Floor-Ceiling Assemblies using the Tapping Machine, ASTM E492, 1996.
59. American Society for Testing and Materials, "Standard

Test Method for Field Measurement of Tapping Machine Impact Sound Transmission through Floor-Ceiling Assemblies and Associated Support Structures," ASTM E1007.

60. International Standardization Organization, "Acoustics, Rating of Sound Insulation in Buildings and of Building Elements, Impact Sound Insulation," ISO 717-2, 1982.
61. American Society for Testing and Materials, "Classification for Determination of Impact Insulation Class (IIC)," ASTM E989, 1994.
62. W. Fasold, "Untersuchungen über den Verlauf der Sollkurve für den Trittschallschutz im Wohnungsbau," *Acustica*, Vol. 15, 1965, pp. 271–284.
63. T. Mariner and W. W. Hehmann, "Impact Noise Rating of Various Floors," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 206–214.
64. R. Josse, "How to Assess the Sound-Reducing Properties of Floors to Impact Noise (Footsteps)," *Appl. Acoust.*, Vol. 5, 1972, pp. 15–20.
65. E. Gerretsen, "A New System for Rating Impact Sound Insulation," *Appl. Acoust.*, Vol. 9, 1976, pp. 247–263.
66. K. Bodlund, "Alternative Reference Curves for Evaluation of the Impact Sound Insulation between Dwellings," *J. Sound Vib.*, 1985, pp. 381–402.
67. A. C. C. Warnock, "Floor Impact Noise and Foot Simulators," *Proceedings Inter-Noise 83*, 1983, p. 1127.
68. A. C. C. Warnock, "Low Frequency Impact Sound Transmission through Floor Systems," *Proceedings Inter-Noise 92*, 1992, p. 743.
69. International Standardization Organization, "Acoustics, Measurement of Sound Insulation in Buildings and of Building Elements, Laboratory Measurement of the Reduction of Transmitted Impact Noise by Floor Covering on a Standard Floor," ISO 140-8, 1978.
70. L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, Berlin, Heidelberg, New York, Springer-Verlag, 1988.
71. E. Sonntag, "Kurzprüfverfahren für das Verbesserungsmaß des Trittschallschutzes bei Weichbelägen, Bauforschung," *Baupraxis*, Vol. 7, 1977, pp. 81–89.
72. International Standardization Organization, "Acoustics, Determination of Dynamic Stiffness, Materials Used under Floating Floors in Dwellings," ISO 9052-1, 1989.
73. E. E. Ungar, "Design of Floated Slabs to Avoid Stiffness of Entrapped Air," *Noise Con. Eng.*, p. 12, July–August, 1975.
74. DIN 4109, Schallschutz im Hochbau, Anforderungen und Nachweise, 1989.
75. VDI 4100, VDI-Richtlinien, Schallschutz von Wohnungen, Kriterien für Planung und Beurteilung, E 1989.
76. R. A. Hedeen, "NIOSH Compendium on Materials for Noise Control, DHEW Publication No. 80-116, U.S. Department of Health, Education and Welfare, Cincinnati, OH.
77. R. B. Dupree, "Catalog of STC and IIC Ratings for Wall and Floor/Ceiling Assemblies," Office of Noise Control, California Department of Health Services, Berkeley, CA.
78. A. C. C. Warnock and D. W. Monk, "Sound Transmission Loss of Masonry Walls: Tests on 90, 140, 190, 240 and 290 mm Concrete Block Walls with Various Surface Finishes," Building Research Note 217, NRCC.
79. T. D. Northwood and D. W. Monk, "Sound Transmission Loss of Masonry Walls: Twelve-Inch Lightweight Concrete Blocks—Comparison of Latex and Plaster Sealers," Building Research Note 93, NRCC.
80. T. D. Northwood and D. W. Monk, "Sound Transmission Loss of Masonry Walls: Twelve-Inch Lightweight Concrete Blocks with Various Surface Finishes," Building Research Note 90, NRCC.

94

RATINGS AND DESCRIPTORS FOR THE BUILDING ACOUSTICAL ENVIRONMENT

GREGORY C. TOCCI

1 INTRODUCTION

Generally, ratings for acoustical environments either set limits on background sound to achieve an acceptable venue or set requirements for audio signal levels to achieve acceptable functionality. In addition, architectural acoustical criteria set limits for sound transmission through building partition constructions, for example, outdoor sound transmitted indoors and indoor sound transmitted between rooms.

This chapter chiefly describes descriptor and rating methods used in establishing acoustical criteria in buildings. Acoustical descriptors characterize in a single number two or more of the three fundamental characteristics of sound: (1) spectral content, (2) level, and (3) time-varying character. Many of the descriptors briefly discussed in this chapter are discussed in more detail elsewhere in this handbook.

The chapter begins with a description of common acoustical descriptors used to quantify sound in the environment. These include sound pressure level, speech interference level, equivalent sound level, and so forth. Many of these are discussed in more detail elsewhere, in particular in Chapter 80. Section 2 of this chapter discusses ratings used to quantify and characterized sound inside and outside of buildings. These include day–night average sound level, noise criteria curves, and so forth. Section 3 discusses descriptors and ratings for the acoustical performance of materials and systems used in the construction of buildings. These generally break down into ratings quantifying the ability of building materials to absorb sound and the ability of building materials and systems to resist the transmission of sound through them. Section 4 discusses descriptors used to assess the quality of sound in building spaces. The discussions draw heavily on standards and methods of the American National Standards Institute (ANSI) and American Society for Testing and Materials (ASTM). Finally, Section 5 discusses additional ratings and descriptors used in International Standards Organization (ISO) standards and standards of selected European and Asian countries.

2 GENERAL ACOUSTICAL DESCRIPTORS

2.1 A-Weighted Sound Pressure Level

The human ear is sensitive to sound extending from approximately 20 to 20,000 Hz but is most sensitive to sound in the 500–4000 Hz frequency range. Above and below this most sensitive frequency range the ear is progressively less sensitive to sound. Since sound measurement instrumentation is typically more uniformly sensitive to sound over a considerably wider frequency range than human hearing, sound level meters use electronic filtering to reproduce the varying sensitivity of the ear to sound at different frequencies. Using this filtering, sound level meters indicate a sound level that is more representative of what the ear hears. This filtering is called A weighting, and the sound pressure levels measured using A-weighted filtering are often called *sound levels* and typically denoted as dB(A) (abbreviation, AL; letter symbol, L_A).

The *A-weighted sound pressure level* is often used to evaluate general human exposure to sound. A-weighted sound pressure levels are most widely used as they are a convenient single-number quantifier of sound over the

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

audible frequency range (as discussed in Chapter 80). Originally, A weighting was intended to represent the varying sensitivity of the ear to sound at sound pressure levels ranging between 40 and 60 dB. Subsequently, B and C weightings were developed to represent the varying sensitivity of the ear to sound over higher sound pressure level ranges. Both ratings allow more low-frequency sound energy with C weighting being nearly flat across the audible frequency range. However, B and C weightings are no longer widely used, with the exception that C weighting is used by some agencies and researchers in assessing transient sound. Also, in the absence of spectral information (octave band sound levels), the difference between A-weighted and C-weighted sound levels can indicate the relative content of low-frequency energy in the sound spectrum.

2.2 Speech Interference Levels

The *speech interference level* (SIL) is the arithmetic average of the sound pressure level in octave bands centered at 500, 1000, 2000, and 4000 Hz.[1] SIL is also discussed in Section 7 of Chapter 80. Because the definitions and octave bands used in the past have changed, ANSI S3.14 recommends that the abbreviation SIL be followed by an indication of the octave bands used for computation, for example, SIL (0.5, 1, 2, 4).

Figure 1 shows the relationship between speech interference level and vocal effort required at various distances to achieve just-reliable speech communication. The crosshatch range indicates *expectant* voice level, that is, the natural inclination of speakers to raise their voices in the presence of high background sound levels. The rate at which the voice is raised ranges between 3 and 6 dB for every 10-dB increase in backaground noise level above 50 dB(A).

The SIL criteria presented in Fig. 1 are applicable to both indoor and outdoor environments and simply relate to speech communication, without any accounting for other circumstances or activities in a space. This figure also includes instructions for its use for determining minimum vocal effort and/or maximum distance to achieve "just reliable" speech articulation (articulation index of 0.45) in the presence of broadband (*flat spectrum*) background noise. The method is not valid for (1) situations where the background sound spectrum departs significantly from being "flat," (2) situations where the background or the signal sound levels vary significantly with time, (3) communications in highly reverberant environments, and (4) situations when speech is distorted.

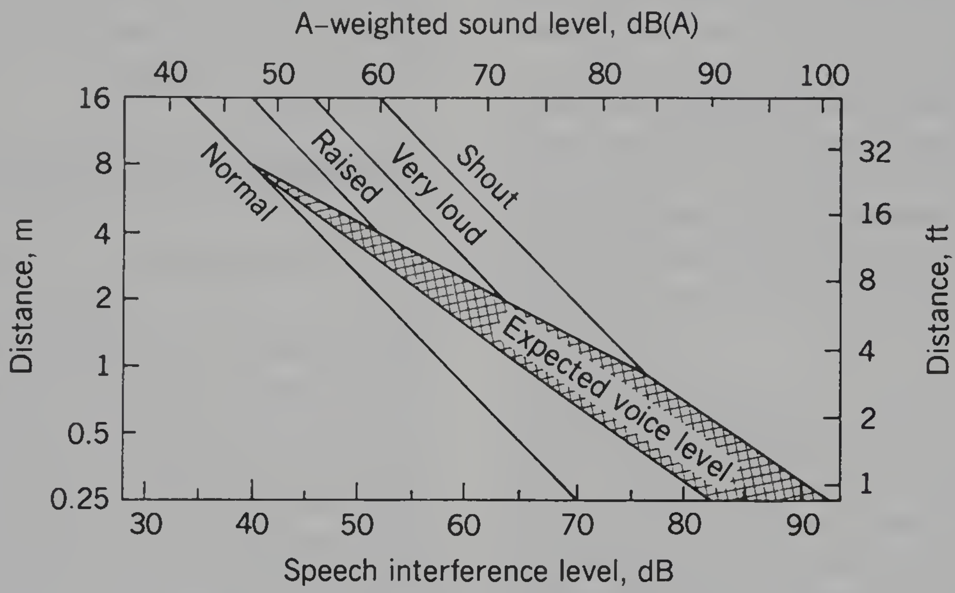


Fig. 1 Talker-to-listener distances for just-reliable communication. The curves show maximum permissible talker-to-listener distances for just-reliable speech communication. The parameter on each curve indicates the relative voice level. Since a talker will raise his or her voice in noise, typically at the rate of 3–6 dB for every 10-dB increase in noise level above 50 dB(A), the expected voice level will increase with increasing noise level. The cross-hatched area shows the range of permissible talker-to-listener distances under these conditions. The lower bound of the cross-hatched area is for voice level being raised at the rate of 3 dB per 10-dB increase in noise level; the upper bound is for a rate of increase of 6 dB per 10-dB increase in noise level. (1) Measure the A-weighted sound level of the background noise using the *slow* response of the sound level meter. (2) Locate this sound level on the upper abscissa. (3) A perpendicular dropped from this point will intersect several curves, indicating the maximum distance between the talker and listener for just-reliable speech communication for various levels of vocal effort.

2.3 Loudness Level

The *loudness level* (LL) is a single-number rating expressed in units of *phons*. It is described more fully in Chapters 80 and 118. The loudness level in phons is the level of a 1000-Hz reference tone with a loudness that appears to be equal to that of the spectrum being rated.[2] The loudness level of an octave band spectrum is obtained by a procedure described in ANSI S3.4-1980(R 1986).[3] The procedure involves overlaying an octave band spectrum of sound to be rated on contours of approximate equal loudness and interpolating *loudness index* values for each octave band level of the spectrum. These loudness index values are then used with specified formulas to arrive at a single-number loudness level rating in phons. Phons are a logarithmic quantity. The corresponding linear quantity is *sones* where 1 sone is the loudness of a sound whose loudness level is 40 phons. A doubling of amplitude in sones corresponds to a factor of 10 increase in loudness level expressed in phons.

2.4 Equivalent Sound Level

As discussed in Chapters 80, 87, and 89, the *equivalent sound level* (QL; symbol, L_{eq}) is the energy average sound level occurring at a particular location over a given time interval. The equivalent sound level is also discussed in Section 9 of Chapter 80. Often, the ambient sound level is monitored for 1-h intervals for environmental sound assessments. When appropriate, hourly equivalent sound levels may be combined in order to determine the equivalent sound level for a 24-h period, or equivalent sound levels for shorter time intervals can be combined to determine the hourly equivalent sound level. Most often, the equivalent sound level is an A-weighted sound pressure level, and the L_{eq} symbol implies A weighting. Some agency criteria incorporate the C-weighted equivalent sound level, symbolized as L_{Ceq}.

2.5 Sound Exposure Level

The *sound exposure level* (SEL; symbols, L_{ET} or L_{AE}) quantifies the total A-weighted acoustic energy integrated over a time interval for a given acoustical event. To simplify the units to decibels, the SEL of an event is described as the hypothetical equivalent sound pressure level enduring for a period of one second that would have the same amount of acoustic energy as the specific transient event for which the SEL was measured. The SEL is used to quantify the sound energy associated with a transient event such as an aircraft over flight. Its practical use requires setting a measurement start and a finish time, but if the maximum sound level occurring during an event is high enough in amplitude compared to levels before and after the event, defining precise start and finish times has little influence on the SEL value reached.

2.6 Single-Event Noise Exposure Level

The *single-event noise exposure level* (SENEL) is similar to the SEL except that it incorporates the integration of sound energy from and to defined points in time before and after a transient event. These points in time are defined as the moments during which the event sound pressure level rises above and falls below a specific sound level threshold. For environmental sound analysis, a commonly used threshold is 65 dB(A). This threshold is not set by any standard, but by measurement personnel seeking a clear point at which the sound pressure level associated with a transient event is sufficiently above the ambient so that lesser transients in the ambient do not trigger a SENEL event. The concept of SENEL has its roots in instrumentation developed to measure transient events from sources such as aircraft and trains.

2.7 Percentile Sound Levels

Percentile sound levels ($L_{XX\%}$) are A-weighted sound pressure levels exceeded an indicated (by the *XX%* value) percent of a time interval. Percentile sound levels are also discussed in Section 11 of Chapter 80. Strictly speaking, the descriptor should also indicate the interval length, for example, L_{90}(1 h). The L_{90}(1 h) is the A-weighted sound pressure level exceeded a total of 54 min out of a continuous 60-min interval. Accordingly, the L_{10}(20 min), is the A-weighted sound pressure level exceeded 2 min out of a continuous 20-min interval.

The L_{10} is representative of intrusive sounds, sounds of short duration, but high level. The L_{90}, on the other hand, is representative of nearly the lowest levels of sound occurring during quiet interludes and is often referred to as the background or residual sound level.

The Federal Highway Administration (FHWA) allows the alternative use of L_{10} or L_{eq} for assessing traffic sound levels and uses these descriptors to set guideline limits for traffic noise exposure at wayside areas according to various categories of land use.[4] The FHWA regulation for assessing the impact of highway noise allows the use of either L_{10} or L_{eq}, but not both.

Some states and municipalities set environmental noise limits based on percentile sound level limits. For example, the Commonwealth of Massachusetts Department of Environmental Protection requires that sound generated by a commercial facility may not cause an ambient sound level that exceeds the preexisting background sound level (L_{90}) without the facility operating by more than 10 dB(A).[5] The City of Cambridge, Massachusetts, requires that construction sound levels not exceed an L_{10}(20 min) of 75 dB(A) at nearest residential property.[6]

3 RATINGS USED TO ASSESS THE ACCEPTABILITY OF AMBIENT SOUND

3.1 Day–Night Average Sound Level

The *day–night average sound level* (DNL; symbol, L_{dn}) is a 24-h average A-weighted sound level where a 10-dB "penalty" is applied to sound occurring between the hours of 10 : 00 p.m. and 7 : 00 a.m. The 10-dB penalty accounts for the heightened sensitivity of a community to noise occurring at night. This descriptor is also discussed in Section 10 of Chapter 80 of named day–night equivalent level. Day–night average sound level has become the primary descriptor for general environmental sound and is often used to assess sound from transportation systems. Among agencies using the day–night average sound level in their criteria and regulations are the U.S. Environmental Protection Agency (EPA),[7] the Federal Aviation Administration (FAA),[8] and the U.S. Department of Housing and Urban Development (HUD).[9]

The U.S. EPA has taken the lead among all federal agencies in unifying usage of environmental sound level descriptors. The EPA has fostered the development of the day–night average sound level but has not enacted regulations controlling general environmental noise. Instead, it has issued guidelines that identify yearly L_{dn} sound levels "sufficient to protect public health and welfare from the effects of environmental noise." Table 1 presents EPA's suggested levels to protect public health and welfare. The EPA specifically cautions that these tabulated levels are not to be used as regulations by other agencies without addressing economic and other considerations associated with sound level restrictions. Of these levels, the most widely cited are a day–night average sound level of 55 dB for outdoor residential areas and a day–night average sound level of 45 dB for indoor residential spaces. Again, these are only to be used as levels with a margin of safety incorporated and not as EPA's recommendations for agency limits.

3.2 Community Noise Equivalent Level

The *community noise equivalent level* (CNEL) is similar to the DNL (L_{dn}) except that, in addition to the 10-dB(A) penalty applied between the hours of 10 : 00 p.m. to 7 : 00 a.m., there is a 5-dB(A) penalty applied to sound between the hours of 7 : 00 p.m. and 10 : 00 p.m. CNEL is defined in Chapter 80. Use of this descriptor seems to be declining nationally, but is most commonly used in California standards, where it is allowed as an alternative environmental sound level descriptor to the day–night average sound level for assessing environmental sound transmission into building spaces.[10,11] It is also used to assess community noise in particular near airports (Chapter 87).

3.3 Noise Criterion Curves

Speech interference level and loudness level are not widely used to evaluate ambient sound in engineering acoustics. More widely used descriptors are noise criterion (NC) curves. Noise criterion curves are a simpler alternative to the evaluation of perceived loudness than the loudness level computation procedure. NC curves are also discussed in Section 8 of Chapter 80. Figure 2 contains the set of NC curves.[12] The NC curves are smoothed versions of the loudness level index curves. The NC value shown for each curve is the

TABLE 1 Yearly L_{dn} Values That Protect Public Health and Welfare with a Margin of Safety

Effect	Level	Area
Hearing	$L_{eq(24)} \le 70$ dB	All areas (at the ear)
Outdoor activity interference and annoyance	$L_{dn} \le 55$ dB	Outdoors in residential areas and farms and other outdoor areas where people spend widely varying amounts of time and other places in which quiet is a basis for use.
	$L_{eq(24)} \le 55$ dB	Outdoor areas where people spend limited amounts of time, such as school yards, playgrounds, etc.
Indoor activity interference and annoyance	$L_{dn} \le 45$ dB	Indoor residential areas
	$L_{eq(24)} \le 45$ dB	Other indoor areas with human activities such as schools, etc.

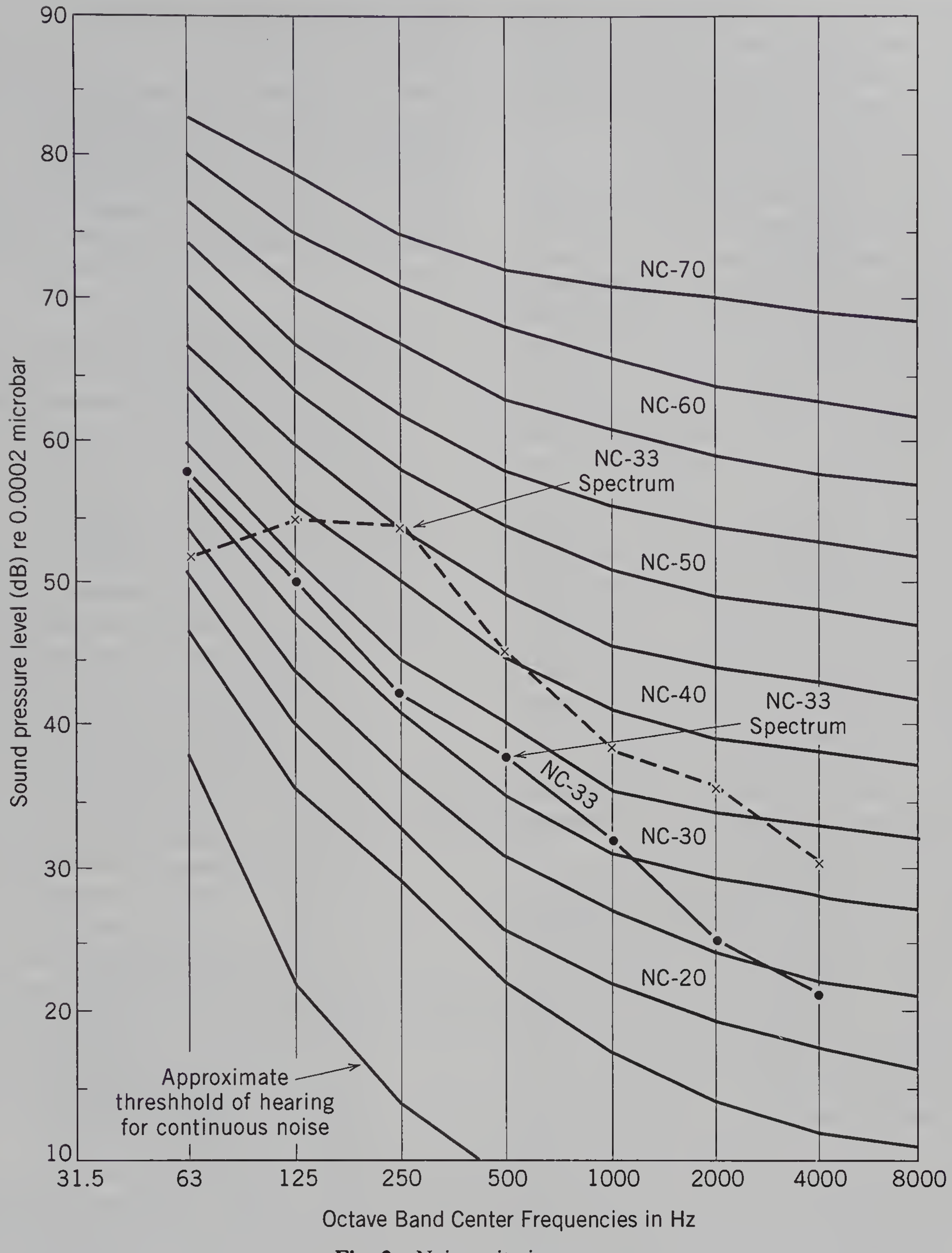


Fig. 2 Noise criterion curves.

SIL (600–1200, 1200–2400, and 2400–4800 Hz), that is, in the "old" octave bands, for each curve. Noise criterion values are commonly determined for an octave band spectrum of sound by superposing the NC curves over the spectrum. The highest NC curve reached by the measured spectrum (tangent to) is the NC rating of the sound pressure level spectrum. Figure 2 shows two typical spectra and the resulting NC values.

Note that the NC curves are defined in terms of the "old" frequency bands 20–75, 75–50, 50–300 cps, and so on. Schultz[13] simply overlaid the "new" preferred octave bands on the NC curves previously defined in old octave bands, added an NC-15 curve, and corrected the approximate threshold of hearing curve originally shown with the NC curves when they were first published. This "updating" of NC curve assessment has not changed the method for evaluating the NC value of a sound pressure spectrum plotted in preferred octave bands.

Often, the NC value of a spectrum is considered within a limited frequency range, such as 500–4000 Hz in the case of concerns for speech intelligibility or 63–125 Hz in the case of low-frequency "rumble"-type sound levels, typically associated with heating, ventilation, and air conditioning (HVAC) systems.

It should also be noted that NC curves are generally used in evaluating continuous sound inside building spaces produced by mechanical systems or environmental sound transmitted into building spaces. Guidelines for assessing the acceptability of sound in buildings generally presume that the background sound does not have tonal or temporal characteristics that lend a distinctive feature to sound, such as transformer hum or fan blade passage tones. Using these criteria to evaluate tonal noise may not be appropriate as it may underestimate the interference of sound on occupant use of spaces.

In the United States background sound in building spaces produced by mechanical systems is not limited by any specific regulation or agency. Instead, the building design profession has, through various organizations, established design criteria for noise in architectural spaces. The most commonly used criteria are the design guidelines for HVAC system noise in unoccupied spaces established by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE). For most of the many years that ASHRAE has published recommended design criteria for spaces, it has expressed them as acceptable ranges of NC curves.

In recent years, ASHRAE has introduced a new set of curves called room criterion (RC) curves, which are discussed below. On their first introduction, ASHRAE publications suggested that RC curves could be used in lieu of NC curves. As the years progressed, the use of RC curves was cited as preferred to the use of NC curves. Finally, the *1995 ASHRAE Applications Handbook* lists recommended ranges of background sound in building spaces expressed only as ranges of RC(N) curves (Ref. 14, p. 43.5). No listing of criteria using NC curves is given in the *1995 Handbook* except that the guidelines state that "if the quality of the sound in the space is of secondary concern, the criteria may be specified in terms of NC levels" (footnote b in Table 2, p. 43.5). Although RC curves represent a better quality of background sound, NC curves are still in wider use at this time as they represent an acceptable compromise between economy and quality of background sound. Table 2 presents the ASHRAE recommended design guidelines for HVAC system noise in unoccupied spaces. These are expressed in RC(N) levels. For use with NC curves, substituting NC levels equal to the RC levels shown is acceptable, subject to the cautionary footnotes in Table 2.

3.4 Room Criterion Curves

ASHRAE has adopted the use of room criterion (RC) curves for evaluating the background sound in building spaces produced by mechanical systems. RC curves are also described in Section 8 of Chapter 80. Unlike NC curves, RC curves are straight lines, as shown in Fig. 3. These curves are used by superposing them on a measured octave band spectrum. The RC value of the spectrum is determined through the following procedure:

1. Determine the arithmetic average, to the nearest whole number, of the sound pressure levels in the 500-, 1000-, and 2000-Hz octave frequency bands. This is the RC level associated with the room sound level spectrum.
2. Draw a line having a −5 dB/octave slope through the RC level at 1000 Hz determined from step 1.
3. Classify the subjective quality or character of the room sound level spectrum as follows:
    a. *Neutral.* A spectrum classified as neutral is free of tonal quality and would be perceived as unobtrusive or bland. A neutral sound spectrum falls exactly along or close to a single RC contour. If the octave band data do not exceed the RC curve determined in step 2 by more than 5 dB at and below 500 Hz and do not exceed the RC curve by more than 3 dB at and above 1000 Hz, the spectrum is considered neutral, and the designator (N) is placed after the RC level.
    b. *Rumble.* A sound spectrum that is perceived to have a "rumbly" quality has an excess of low-frequency sound energy. A rumbly spectrum is characterized as one with octave band sound levels that exceed the RC curve determined in step 2 by more than 5 dB at and below 500 Hz. For such spectra the designator (R) is placed after the RC level.
    c. *Hiss.* A sound spectrum that is perceived to have a "hissy" quality has an excess of high-frequency energy. A hissy spectrum is characterized as one with octave band sound levels that exceed the RC curve determined in step 2 by more than 3 dB above 500 dB. For such spectra the designator (H) is placed after the RC level.
    d. *Acoustically Induced Perceptible Vibration.* The cross-hatched region of the RC curves in Fig. 3 indicates sound pressure levels in the 16- to 63-Hz octave bands at which perceptible vibration in building walls and ceilings can occur. These sound levels often produce rattles in cabinets, doors, pictures. lighting fixtures, and so forth.

TABLE 2 Design Guidelines for HVAC System Noise in Unoccupied Spaces

Space	RC(N) Level[a, b]
Private residences, apartments, condominiums	25–35
Hotels/Motels	
Individual rooms or suites	25–35
Meeting/banquet rooms	25–35
Halls, corridors, lobbies	35–45
Service/support areas	35–45
Office buildings	
Executive and private offices	25–35
Conference rooms	25–35
Teleconference rooms	25 (max)
Open plan offices	30–40
Circulation and public lobbies	40–45
Hospitals and clinics	
Private rooms	25–35
Wards	30–40
Operating rooms	25–35
Corridors	30–40
Public areas	30–40
Performing arts spaces	
Drama theaters	25 (max)
Concert and recital halls	—[c]
Music teaching studios	25 (max)
Music practice rooms	35 (max)
Laboratories (with fume hoods)	
Testing/research, minimal speech communication	45–55
Research, extensive telephone use, speech communication	40–50
Group teaching	35–45
Churches, mosques, synagogues	25–35
With critical music programs	—[c]
Schools	
Classrooms up to 750 ft^2	40 (max)
Classrooms over 750 ft^2	35 (max)
Lecture rooms for more than 50 (unamplified speech)	35 (max)
Libraries	30–40
Courtrooms	
Unamplified speech	25–35
Amplified speech	30–40
Indoor stadiums and gymnasiums	
School and college gymnasiums and natatoriums	40–50[d]
Large seating capacity spaces (with amplified speech)	45–55[d]

---

[a] The values and ranges are based on judgment and experience, not on quantitative evaluations of human reactions. They represent general limits of acceptability for typical building occupancies. Higher or lower values may be appropriate and should be based on a careful analysis of economics, space usage, and user needs. They are not intended to serve by themselves as a basis for a contractual requirement.

[b] When the quality of sound in the space is important, specify criteria in terms of RC(N). If the quality of the sound in the space is of secondary concern, the criteria may be specified in terms of NC levels.

[c] An experienced acoustical consultant should be retained for guidance on acoustically critical spaces (below RC 30) and for all performing arts spaces.

[d] Spectrum levels and sound quality are of lesser importance in these spaces than overall sound levels.

(Reprinted with permission from the *1995 ASHRAE Handbook—HVAC Applications*.)

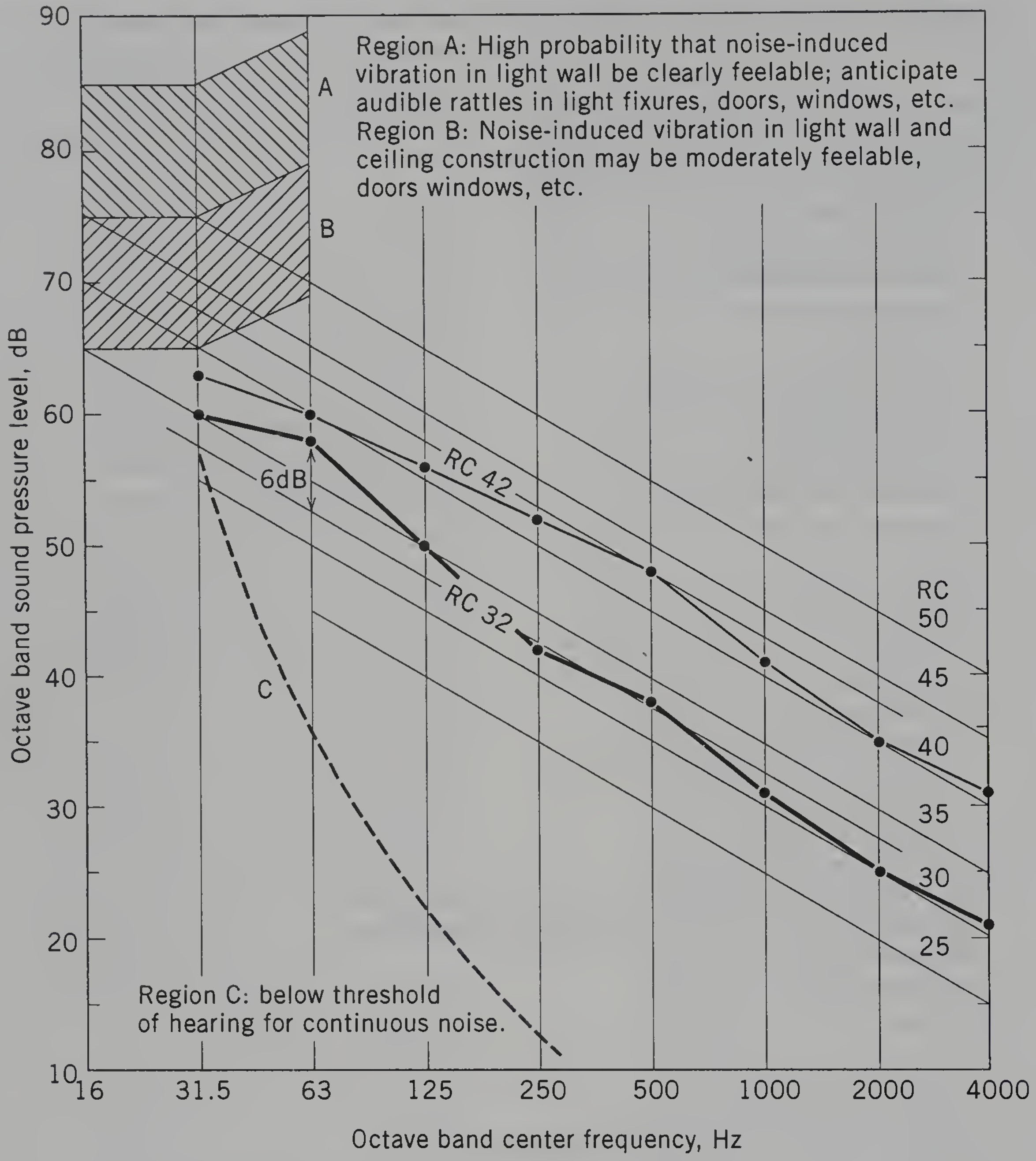


Fig. 3 Room criterion curves plus examples of RC 42(N) and RC 32(R) sound level spectra. (Reprinted with permission from the *1995 ASHRAE Handbook—HVAC Applications*.

For spectra with levels that fall into this range, the designator (RV) is placed after the RC level.

Figure 3 also provides examples of two types of spectra. In one case, the spectrum follows along the RC 42 curve (i.e., the average of the sound pressures in the 500-, 1000-, and 2000-Hz octave bands is 42 dB). In this spectrum, all of the octave band levels are within the prescribed limits, and is therefore designated an RC 42(N) spectrum. The spectrum labeled RC 32(R) is generally 10 dB lower than the RC 42(N) spectrum, except in the low frequencies where the 63-Hz octave band exceeds the RC 32 contour by 6 dB, thus constituting a low-frequency rumble component.

3.5 Balanced Noise Criterion Curves

Beranek[15,16] has drawn upon low-frequency hearing considerations to extend NC curves down to low frequencies. He has also sloped downward the high-frequency ends of the NC curves to reduce the subjectively hissy nature of sound spectra that conform to NC spectra shapes, thereby creating a "balanced" spectrum subjectively perceived as more uniform and devoid of significant tonal content. A set of balanced noise criterion curves (NCBs) are presented in Fig. 4. NCB curves are also discussed in Section 8 of Chapter 80. As with RC curves, NCB curves are accompanied by a procedure for assessing the perceived balance of a sound spectrum, that is, whether or not a spectrum will be perceived as rumbly

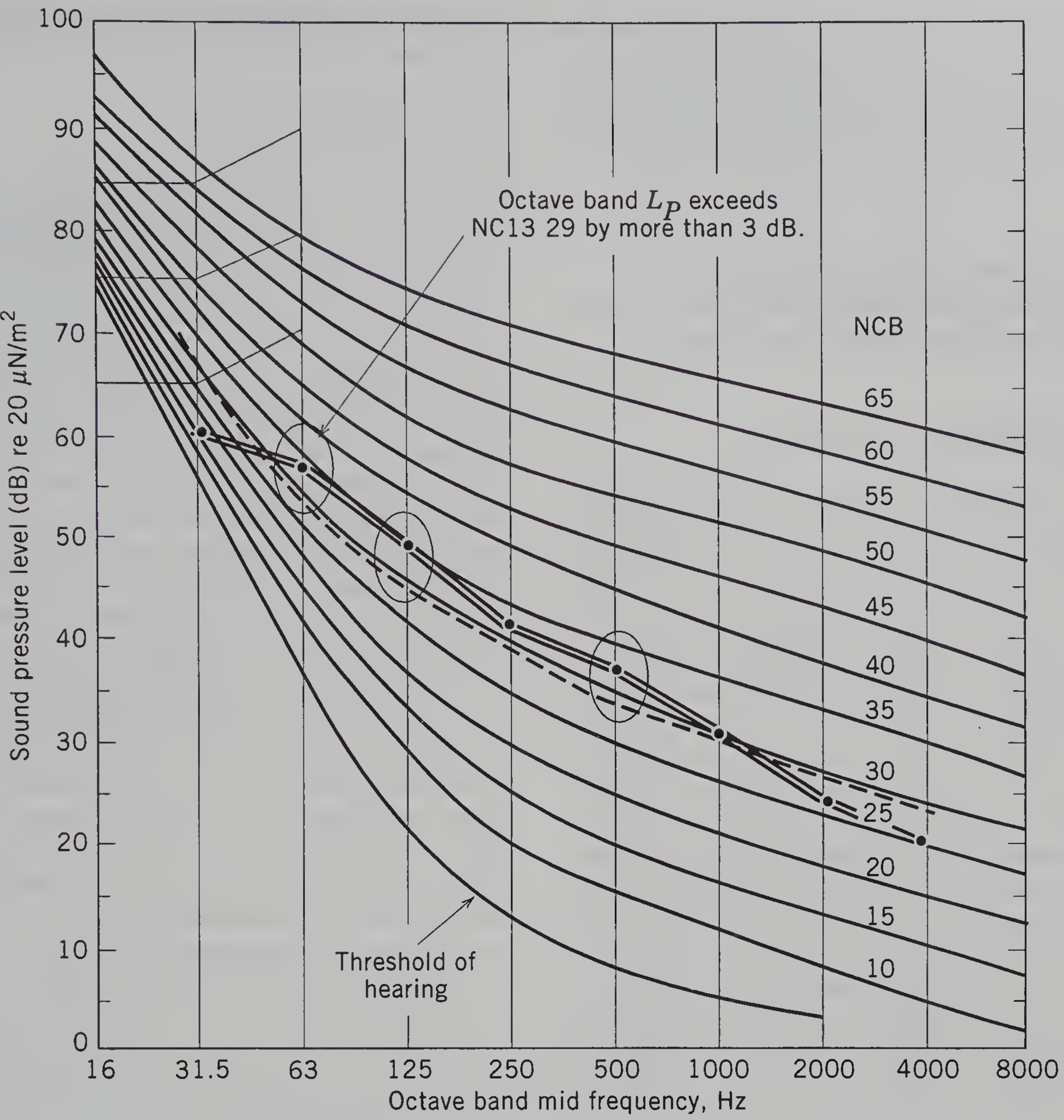


Fig. 4 Balanced noise criterion curves for occupied rooms. Octave-band sound pressure levels of the magnitudes indicated in regions *A* and *B* may induce audible rattles or feelable vibrations in lightweight partitions and ceiling constructions (e.g., thin plaster or gypsum board on metal framing) as follows (taken from Ref. 12): region *A*, clearly noticeable vibrations; region B, moderately noticeable vibrations.

or hissy, and so forth. This procedure is much like that used with the RC curves, but designators (N), (R), (H), and (RV) are not used as part of the descriptor as is the case with RC values. Rules for assessing rumble and hissy qualities of sound spectra using NCB curves are as follows:

1. Determine the L_{SIL} for the spectrum being evaluated. The L_{SIL} is the arithmetic average of sound levels in the 500-, 1000-, 2000-, and 4000-Hz octave bands rounded to the nearest decibel, for example, XX. This spectrum would then be denoted an NCB-XX spectrum.
2. The spectrum will be perceived as rumbly if any octave band sound level at or below 1000 Hz is above the NCB-YY curve. YY is equal to the XX value in step 1 plus 3 dB.
3. The spectrum will be perceived as hissy if any octave band sound level at frequencies above 500 Hz exceeds the NCB-ZZ curve. To determine the value ZZ, first determine the arithmetic average of sound pressure levels in the three octave bands 125 through 500 Hz. Then determine which NCB curve has this sound pressure level at 250 Hz. This is the NCB-ZZ curve.

Figure 4 also illustrates the evaluation of a spectrum using NCB curves. The example spectrum is equal to the lower of the two spectra shown in Fig. 3 used to illustrate the use of RC curves. Note that using the NCB curves, this spectrum exhibits rumbly characteristics in the 63-, 125-, and 500-Hz octave bands where octave band sound levels exceed the NCB 29 curve by more than 3 dB.

4 DESCRIPTORS AND RATINGS FOR MATERIALS AND BUILDING CONSTRUCTIONS THAT AFFECT SOUND INSIDE BUILDING SPACES

4.1 Sound Absorption Coefficient

The ability of a material to absorb sound is characterized by the sound absorption coefficient (α). This is the ratio of the amount of sound energy absorbed to the amount incident on the surface of a sound-absorbing material. Theoretically, this can range between 0 and 1.0. The higher the sound absorption coefficient the better the ability of the material to absorb sound. Normally, materials considered to be sound absorbing have sound absorption coefficients greater than or equal to 0.5 (i.e., 50%).

Because sound-absorbing materials have differing sound absorption performances at different frequencies, sound absorption coefficients are generally determined via measurements and reported in octave band frequencies—generally 125, 250, 500, 1000, 2000, and 4000 Hz. Sound absorption coefficient is measured in a reverberation chamber[17] using any one of a variety of specified mounting methods.[18] Mounting methods, especially clearance between the material and hard backing surface (reverberation room walls, floor, or ceiling) can affect the measured sound absorption coefficient significantly.

4.2 Noise Reduction Coefficient

To simplify comparison between sound-absorbing materials, a single-value descriptor called the noise reduction coefficient (NRC)[17] is used. It is the arithmetic average of the sound absorption coefficients at 250, 500, 1000, and 2000 Hz. Generally, for sound-absorbing materials, the NRC falls between 0.55 and 0.95. Again, the higher the NRC, the higher the sound absorption performance of a material. Table 3 also presents noise reduction coefficients for the common building materials listed.

Often, the sound absorption of building components, such as upholstered chairs in an auditorium or room sound absorption devices sold as units, are quantified in sabins, either metric (square metres) or English (square feet). This is the case for chairs, pews, and seated audience in Table 3. One sabin is an amount of room absorption equivalent to one square foot (or one square metre in metric units) of material having a sound absorption coefficient of 1.0.

4.3 Noise Reduction

Rather than specifying a specific sound level limit inside building spaces, standards and regulations set by government agencies and organizations often control interior sound levels by specifying minimum required noise reductions (NRs) of exterior facades or interior building partitions. Noise reduction is defined as the difference between the sound pressure level produced by a sound source on one side (the source side) and the sound pressure level on the other side (the receiver side) of a construction being evaluated. Noise reduction is most often measured in octave bands, but sometimes it is expressed as an A-weighted sound pressure level difference. Note that the A-weighted noise reduction depends on the source spectrum. For example, the A-weighted noise reduction for a building exterior wall exposed to aircraft sound may be different from the A-weighted noise reduction for traffic sound. This is because aircraft and traffic sound have their respective energies concentrated in different frequency ranges, and the sound isolation performance for a wall in those two ranges may differ. There are no widely available standards for measuring noise reduction; however, ASTM has formalized the procedure for measuring the noise reduction between interior spaces in buildings in ASTME597-81 (R1987).[19]

Noise reduction is not only related to the ability of a wall construction to resist the transmission of sound, but it is also related to the size of the transmitting wall and the reverberant character of the receiver room. Hence, a wall construction used in different circumstances may result in different noise reductions. Most importantly, noise reduction is significantly affected by the quality of construction, specifically with respect to sound leakage paths associated with gaps and penetrations in the construction, including the perimeter of doors and windows, electrical outlets, fixtures, ductwork, and so forth.

4.4 Sound Transmission Loss

Sound transmission loss (TL) is 10 times the logarithm of the ratio between the sound energy incident on a material or sound isolation construction, divided by the sound energy transmitted through this material. Unlike noise reduction, sound transmission loss is measured in a laboratory according to standards established in ASTM E 90.[20] The standard requires that sound transmission loss be measured in one-third octave frequency bands in a source room–receiver room test suite that complies with certain standards of sound isolation and room dimensions. The result of the test standard is a one-third octave

TABLE 3 Coefficients of General Building Materials and Furnishings

Material	Sound Absorption Coefficient (α)						
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	NRC
Brick, unglazed	0.03	0.03	0.03	0.04	0.05	0.07	0.05
Carpet, heavy, on concrete	0.02	0.06	0.14	0.37	0.60	0.65	0.45
Same, on 40-oz hairfelt or foam rubber	0.08	0.24	0.57	0.69	0.71	0.73	0.55
Concrete block, painted	0.10	0.05	0.06	0.07	0.09	0.08	0.05
Fabrics							
Light velour, 10 oz/yd^2, hung straight, in contact with wall	0.03	0.04	0.11	0.17	0.24	0.35	0.15
Medium velour, 14 oz/yd^2, draped to half area	0.07	0.31	0.49	0.75	0.70	0.60	0.55
Heavy velour, 18 oz/yd^2, draped to half area	0.14	0.35	0.55	0.72	0.70	0.65	0.60
Floors							
Concrete or terrazzo	0.01	0.01	0.015	0.02	0.02	0.02	0
Linoleum, asphalt, rubber or cork tile on concrete	0.02	0.03	0.03	0.03	0.03	0.02	0.05
Wood	0.15	0.11	0.10	0.07	0.06	0.07	0.1
Glass, large panes of heavy plate	0.18	0.05	0.04	0.03	0.02	0.02	0.05
Gypsum board, $\frac{1}{2}$ in. nailed to 2 × 4's 16 in. on center	0.29	0.10	0.05	0.04	0.07	0.09	0.05
Marble or glazed tile	0.01	0.01	0.01	0.01	0.02	0.02	0
Openings							
Stage, depending on furnishings				0.25–0.75			
Deep balcony, upholstered seats				0.50–1.00			
Grilles, ventilating				0.15–0.50			
Plaster, gypsum or lime, smooth finish on tile or brick	0.13	0.15	0.02	0.03	0.04	0.05	0.05
Plaster, gypsum or lime, rough finish on lath	0.02	0.03	0.04	0.05	0.04	0.03	0.05
Same, with smooth finish	0.02	0.02	0.03	0.04	0.04	0.03	0.05
Plywood paneling, $\frac{7}{8}$ in. thick	0.28	0.22	0.17	0.09	0.10	0.11	0.05
Water surface, as in a swimming pool	0.008	0.008	0.013	0.015	0.020	0.025	0
Air, sabins per 1000 ft^3	—	—	—	—	2.3	7.2	—

Type of Seats/Audience	Absorption[a]					
	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Audience, seated in upholstered seats, ft^{-2} of floor area	0.60	0.74	0.88	0.96	0.93	0.85
Unoccupied cloth-covered upholstered seats, ft^{-2} of floor area	0.49	0.66	0.80	0.88	0.82	0.70
Unoccupied leather-covered upholstered seats, ft^{-2} of floor area	0.44	0.54	0.60	0.62	0.58	0.50
Wooden pews, occupied, ft^{-2} of floor area	0.57	0.61	0.75	0.86	0.91	0.86
Chairs, metal or wood seats, each, unoccupied	0.15	0.19	0.22	0.39	0.38	0.30

[a] Values given are in sabins per square foot of seating area or per unit.

band sound transmission loss spectrum extending from 100 (80 optionally) to 5000 Hz.

Unlike noise reduction, the TL is influenced only by mass, stiffness, and damping; and how a construction is configured. Noise reduction is influenced equally by these but also by the surface area size of the partition and the acoustical conditions of the source and receiver rooms. Specifying the performance of a wall construction is awkward because NR is most often influenced by reverberant room characteristics and/or sound flanking conditions that are not part of the wall. Specifying the performance of a wall construction in terms of TL is more logical in that TL is only influenced by characteristics directly related to properties of the material or wall.

4.5 Sound Transmission Class Rating

Since sound TL is frequency dependent, it is generally reported in the third octave frequency bands between 125 and 4000 Hz. As a convenience, a single-number rating method has been developed that allows a single value to be ascribed to a *transmission loss spectrum* that is the set of 16 one-third octave band TL values. This rating is referred to as the sound transmission class (STC), which has been defined in the ASTM Standard E413.[21] This standard defines a procedure for determining the STC value for a TL spectrum that involves fitting a specified contour to one-third octave band TL data.

The STC value was originally developed to approximate the performance of a material in reducing the transmission of intelligible speech sounds. The STC characterization of TL is useful for a quick comparison of materials, but it is not necessarily useful for assessing the performance of a material with respect to other nonspeech sounds such as music and transportation systems sources such as aircraft, highway traffic, train passbys, and so forth. For this purpose, it is necessary to consider octave or third octave band TL spectra with specific attention, particularly, to the low-frequency range, typically at frequencies below 500 Hz. The STC contour is shown in Fig. 5. Determining the STC for a partition construction and/or material involves fitting this contour to the 16 one-third octave band TL values. This procedure involves raising or lowering the contour until the following two rules are met:

1. The contour may not be raised above the point at which the TL in any one-third octave band falls more than 8 dB below the contour.
2. The contour may not be raised above the point at which the total number of deficiencies is greater than 32.

A deficiency occurs when the TL data in any one one-third octave band falls below the contour. For example, three deficiencies would be when the TL data in any one one-third octave fall 3 dB below the contour, or when the data in three one-third octave bands falls 1 dB below the contour. Figure 5 further illustrates the meaning of deficiency.

The STC rating from the contour fitting procedure is the TL value of the contour at 500 Hz. This is illustrated in Figs. 5 and 6.

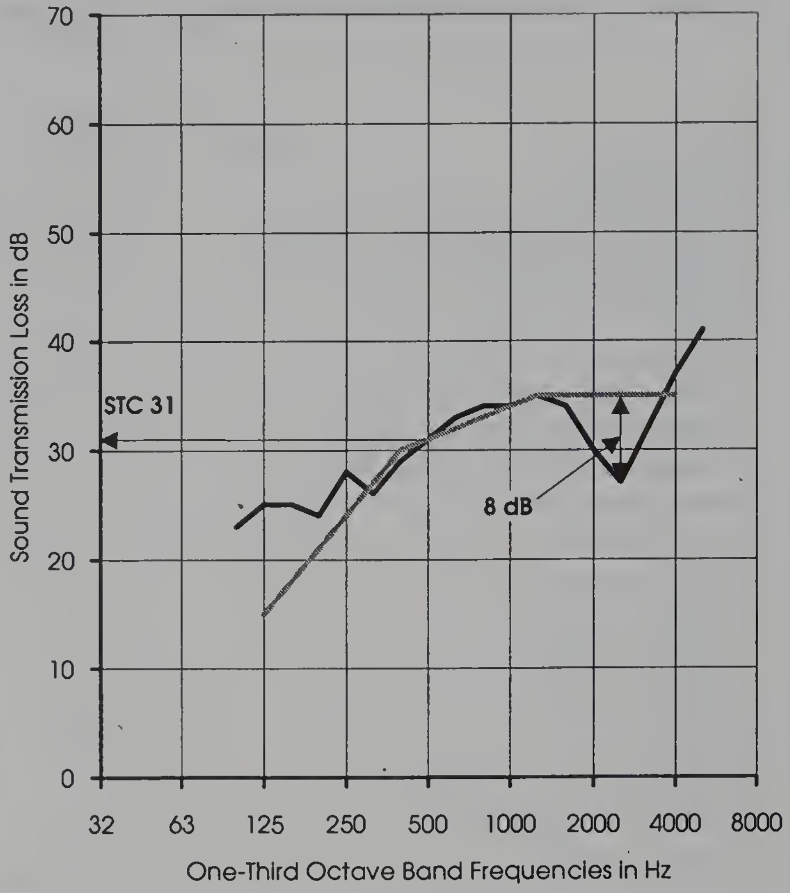


Fig. 5 A $\frac{1}{4}$-in. glass sound transmission loss and STC contour.

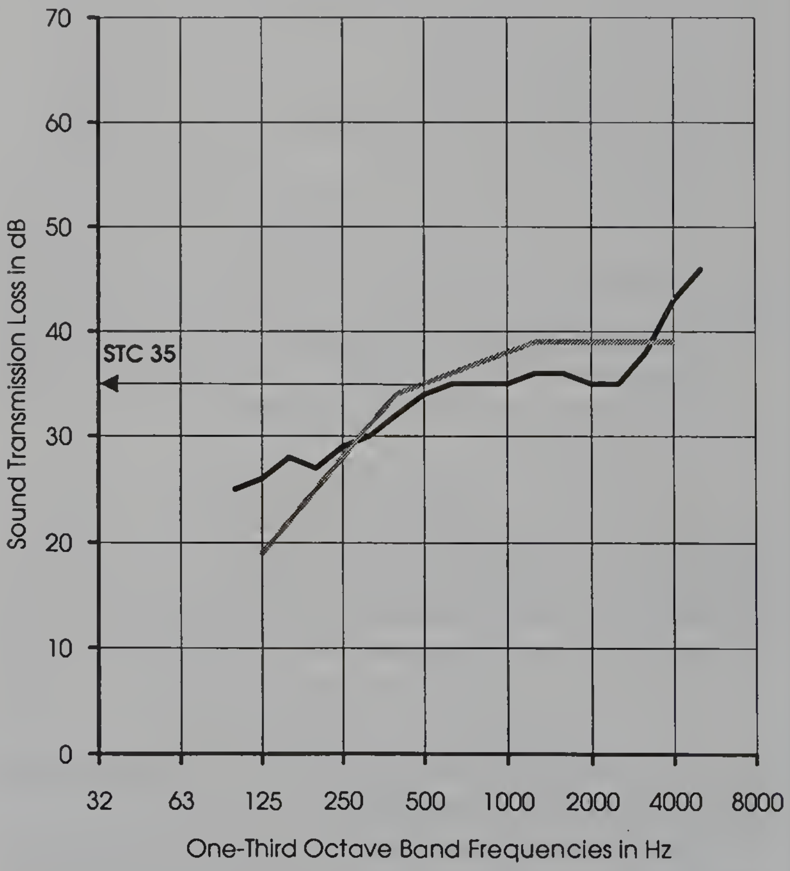


Fig. 6 A $\frac{1}{4}$-in. laminated glass sound transmission loss and STC contour.

In Fig. 5, $\frac{1}{4}$-in. monolithic glass is shown to have an STC rating of 31. In this example, the STC contour placement is constrained by the maximum allowed 8 dB deficiency at 2500 Hz (rule 1).

In Fig. 6, which shows TL data for $\frac{1}{4}$-in. laminated glass, the dip in the TL data, characteristic of $\frac{1}{4}$-in. monolithic glass, is removed by the damping interlayer. In the case of $\frac{1}{4}$-in. laminated glass, the STC contour placement is constrained by the maximum 32-dB deficiency requirement (rule 2).

4.6 Field Sound Transmission Class Rating

The field sound transmission class (FSTC) rating is used for in situ wall and floor/ceiling sound isolation performance assessment.[22] The standard requires the measurement of sound transmission loss and includes a required procedure to show that the FSTC rating, as it has been determined by the test procedure, was not influenced by flanking of sound around the partition intended to be tested. Sound transmission class and FSTC ratings are intended by standard to be equivalent; however, practical experience indicates that FSTC ratings tend to be up to five rating points less than laboratory-measured STC ratings.

4.7 Noise Isolation Class

The noise isolation class (NIC) rating is similar to the STC and FSTC. However, the standard STC rating contour is applied to the one-third octave band noise reduction data measured in a field situation, rather than the transmission losses measured in the field.[21] No correction to the measured noise reduction data is made to account for partition size, receiving room absorption, or sound flanking. Like the STC and FSTC ratings, the field measured NIC rating of a noise reduction spectrum is equal to the value of the contour crossing at 500 Hz. In the absences of sound flanking, the NIC is generally within five points of the laboratory STC rating for typical building partition constructions. The NIC rating is used to assess the sound isolation performance of in situ partition constructions, especially complicated ones that involve multiple sound transmission paths that are not suited for laboratory testing. There are no widely used standards using the NIC rating; however, the NIC rating is often used in lieu of STC and FSTC ratings.

4.8 Normalized Noise Isolation Class

The normalized noise isolation class (NNIC) is the same as the NIC rating except the receiving room absorption is normalized to correspond to a 0.5-s reverberation time.[21]

4.9 Outdoor–Indoor Transmission Class

The STC rating was developed to assess the sound isolation performance of partitions and materials with respect to speech and household sounds, that is, sounds produced by home appliances and entertainment equipment.[23] “Household sound” tends to be rich in mid- and high-frequency sound energy. However, environmental sound, typically associated with transportation systems and equipment, is rich in low- and mid-frequency sound energy. The outdoor–indoor transmission class (OITC) rating is a single-number rating that can be used for comparing the sound isolation performance of building facades and facade elements. The rating has been devised to quantify the ability of these to reduce the perceived loudness of ground and air transportation noise transmitted into buildings. This standard is contained in ASTM E 1332-90 Standard Classification for Determination of Outdoor–Indoor Transmission Class.

The standard establishes a single-number rating, the OITC, by defining a standard spectrum of ground and air transportation noise. This spectrum is used with sound transmission loss data measured in a laboratory using the ASTM E 90 method and a mathematical relationship given in the standard to calculate the OITC rating.

The OITC rating is similar to the STC rating in that it uses ASTM E 90 TL data and uses this data to derive a single-number rating that increases with increasing sound isolation ability. It differs in that the OITC does not involve a contour fitting process but instead uses a standard spectrum and a mathematical relationship. The mathematical relationship is as follows:

$$\text{OITC} = 100.14 - 10 \log \sum_{f} 10^{(L_f - TL_f + A_f)/10} \quad \text{dB}.$$

Table 4 presents a worksheet summarizing the method for calculating the OITC rating for $\frac{1}{4}$-in. glass. As indicated in Table 2, the OITC rating for $\frac{1}{4}$-in. glass is 29. Note that the TL in the 80-Hz band in the worksheet of Table 4 is not shown since it was not measured. The actual TL value in the 80-Hz band, were it included, would not affect the OITC rating indicated in this table.

4.10 Impact Isolation

In residential buildings, quantifying the impact isolation of floor/ceiling constructions with respect to footfall is defined in ASTM E 492-90[25] and ASTM E 1007-90[26] for laboratory and field situations, respectively. Both standards involve placing a mechanical impacting device called a tapping machine on the floor of the construction to be tested and measuring resulting sound levels below.

TABLE 4 Worksheet for Calculating the OITC Rating of $\frac{1}{4}$-in. Glass

Band Center Frequency, Hz (1)	Reference Sound Spectrum, dB (2)	A-Weighting Correction, dB (3)	Column 2 plus Column 3, dB (4)	Specimen TL, dB (5)	Column 4 minus Column 5, dB (6)	$10^{(\text{Column } 6/10)}$
80	103	−22.5	80.5			
100	102	−19.1	82.9	23	59.9	977,237
125	101	−16.1	84.9	25	59.9	977,237
160	98	−13.4	84.6	25	59.6	912,011
200	97	−10.9	86.1	24	62.1	1,621,810
250	95	−8.5	86.4	28	58.4	691,831
315	94	−6.6	87.4	26	61.4	1,380,384
400	93	−4.8	88.2	29	59.2	831,764
500	93	−3.2	89.8	31	58.8	758,578
630	91	−1.9	89.1	33	56.1	407,380
800	90	−0.8	89.2	34	55.2	331,131
1000	89	0.0	89.0	34	55.0	316,228
1250	89	0.6	89.6	35	54.6	288,403
1600	88	1.0	89.0	34	55.0	316,228
2000	88	1.2	89.2	30	59.2	831,764
2500	87	1.3	88.3	27	61.3	1,348,963
3150	85	1.2	86.2	32	54.2	263,027
4000	84	1.0	85.0	37	48.0	63,096
						12,317,071

10 log(Sum) = 70.91

OITC = 100.14 − 10 log(Sum) = 29

From Ref. 24.

The design and use of the tapping machine are described in ISO 140/6.[27] The result of the test is a one-third octave band impact sound pressure level in the room beneath the floor/ceiling construction being tested.

4.11 Impact Insulation Class

Similar to the STC, the impact insulation class (IIC) rating is a single-number rating that facilitates the comparison of the impact sound isolation performance of floor/ceiling constructions. The IIC rating is determined by fitting a standard contour to the one-third octave band sound pressure level data measured in accordance with ASTM E 492.

4.12 Field Impact Insulation Class

The field impact insulation class (FIIC) rating is the same as the IIC rating except it is used to rate the impact sound insulation performance of in situ floor/ceiling constructions and is used in conjunction with the ASTM E 1007 test method. The standard includes procedures for assessing flanking conditions.

5 DESCRIPTORS ASSESSING THE QUALITY OF SOUND INSIDE BUILDING SPACES

5.1 Reverberation Time

The reverberation time (T_{60}) is the time it takes for sound to decay by 60 dB(A). Reverberation time is controlled in interior building spaces to enhance speech intelligibility or enhance music quality. Generally, speech intelligibility benefits from a short reverberation time (typically 1 s) while orchestral concert and organ music benefit from long reverberation times (2–5 s). Figure 7 presents criteria for reverberation time in spaces according to use for music or speech and according to room volume.

There are four generally accepted methods for calculating the reverberation time. The first and probably most widely used is the *Sabine equation*. It is more appropriately used in moderately to highly reverberant spaces. The Sabine equation is as follows:

$$T_{60} = \frac{0.049V}{A} \quad \text{(English units).}$$

The second method for determining room reverbera-

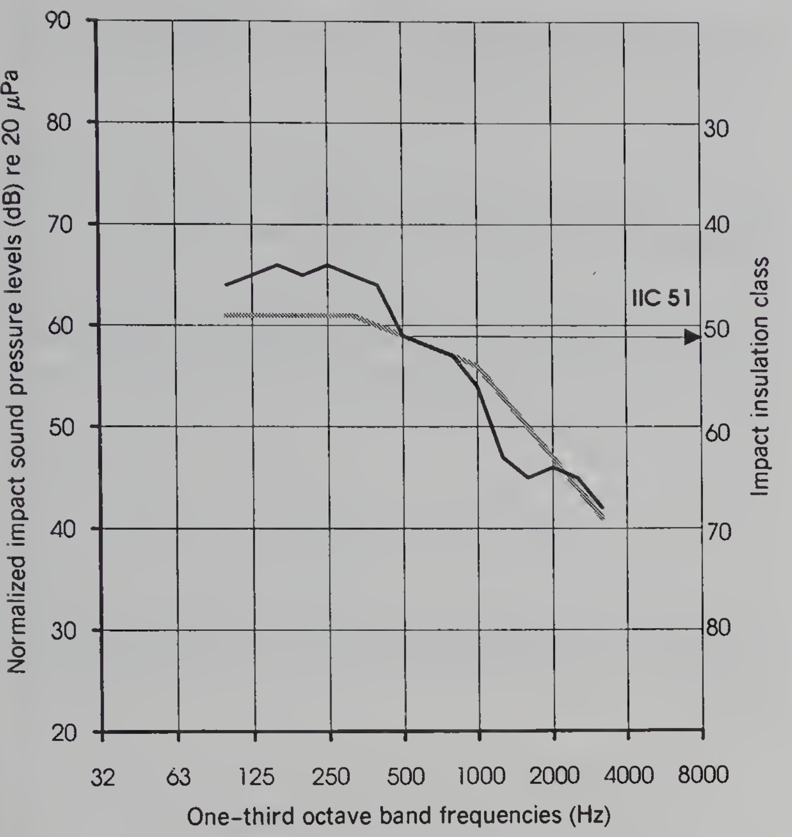


Fig. 7 Typical impact sound level spectrum and IIC contour.

tion is the *Norris Eyring equation.* It is most appropriately used in highly absorptive rooms. The Norris Eyring equation is as follows:

$$T_{60} = \frac{0.049V}{(-S)\ln(1-\overline{\alpha})} \quad \text{(English units.}$$

The third method for determining room reverberation is the *Fitzroy equation.* It is used primarily for rooms where the absorption of parallel surfaces is approximately the same. The Fitzroy equation is as follows:

$$T_{60} = \frac{X}{S}\left(\frac{0.049V}{S\overline{\alpha}_x}\right) + \frac{Y}{S}\left(\frac{0.049V}{S\overline{\alpha}_y}\right) + \frac{Z}{S}\left(\frac{0.049V}{S\overline{\alpha}_z}\right) \quad \text{(English units).}$$

In the above equations, V is room volume, S is room surface area, A is room absorption, $\overline{\alpha}$ is average sound absorption coefficient (the x, y, and z subscripts are the average sound absorption coefficients of walls normal to the x-, y-, and z-axes), and X, Y, and Z are the surface areas of walls normal to the x-, y-, and z-axes. If units of volume, surface area, and absorption are metric, then the 0.049 constant in each equation should be 0.161.[28]

5.2 Room Constant

Room constant (R) is defined as the ratio of room sound absorption to the quantity 1 minus the average room absorption coefficient. Using the above symbol definitions, the relationship is expressed as

$$R = \frac{A}{1-\overline{\alpha}}.$$

There are no criteria that make use of room constant; however, it can be useful in assessing the *critical distance*, defined as the distance from a source at which the reverberant field sound pressure level equals the directly radiated sound pressure level.

5.3 Articulation Index

The articulation index (AI) is a weighted proportion of a speech signal usable to convey information. The procedure[29] for estimating articulation index involves accounting for signal strength, background or masking noise, time-varying character of background noise, speech signal clipping, and reverberation. The estimation procedure is carried-out in one-third or full octave frequency bands. The calculated articulation index is used to estimate percent of syllables, words, or sentences understood correctly. This analysis can be used to assess both speech intelligibility or its inverse, speech privacy. AI is also described in Section 6 of Chapter 80.

5.4 Speech Transmission Index

The speech transmission index (STI) is similar to the articulation index. However, the STI uses a modulation transfer function[30,31] in its evaluation and incorporates a procedure for expressing the loss in articulation produced by room reverberation as an equivalent background noise contribution, which can be added to the actual measured background noise level.[32]

5.5 Rapid Speech Transmission Index

The rapid speech transmission index (RASTI) is similar to the STI, uses fewer modulation frequencies, and incorporates only speech and background sound levels in only two octave bands—500 and 2000 Hz. The RASTI value is considered sufficiently accurate only if no nonlinearities in speech transmission exist and if the background sound level is devoid of strong tonal components.

6 DESCRIPTORS AND RATINGS USED IN ISO AND IN SELECTED EUROPEAN AND ASIAN STANDARDS AND REGULATIONS

The definitions of acoustical terms discussed below are based on ISO 140 and ISO 717 standards. Most European and Asian standards and regulations relating to acoustics are based on ISO terminology and methods defined below. Further information regarding the use of these criteria and descriptors is contained in Chapter 97.

6.1 Average Sound Pressure Level in a Room

Ten times the common logarithm of the ratio of the space and time average of the sound pressure squared, to the square of the reference sound pressure.[33] The spatial average is over the entire room with the exception of those areas of the room significantly influenced by direct radiation of a sound source or the near field of the boundaries (wall, etc.). This quantity is denoted by L in decibels:

$$L = 10 \log \frac{p_1^2 + p_2^2 + \cdots + p_n^2}{np_0^2} \quad \text{dB},$$

where $p_1, p_2, \ldots, p_n$ are the root-mean-square (rms) sound pressures at n different positions in the room and $p_0 = 20\ \mu\text{Pa}$ is the reference sound pressure.

6.2 Sound Reduction Index and Transmission Loss

Ten times the common logarithm of the ratio of the sound power W_1 incident on a test specimen to the sound power W_2 transmitted through the specimen.[33] This quantity is denoted by R:

$$R = 10 \log \frac{W_1}{W_2} \quad \text{dB}.$$

The sound reduction index depends on the angle of incidence. If the sound fields are diffuse and if the sound is transmitted only through the specimen, the sound reduction index for diffuse incidence may be evaluated from

$$R = L_1 - L_2 + 10 \log \frac{S}{A} \quad \text{dB},$$

where L_1 is the average sound pressure level in the source room, L_2 is the average sound pressure level in the receiving room, S is the area of the test specimen, which is normally equal to the free test opening, and A is the equivalent absorption area in the receiving room. (*Note:* If the sound fields are not completely diffuse, this equation is an approximation.)

6.3 Apparent Sound Reduction Index and Apparent Transmission Loss

Ten times the common logarithm of the ratio of the sound power W_1 incident on a partition under test to the total sound power W_3 transmitted into the receiving room.[33] This quantity is denoted by R':

$$R' = 10 \log \frac{W_1}{W_3} \quad \text{dB}.$$

In general, the sound power transmitted into the receiving room consists of the sum of the following components:

W_{Dd}, which has entered the partition directly and is radiated from it directly;

W_{Df}, which has entered the partition directly but is radiated from flanking constructions;

W_{Fd}, which has entered flanking construction and is radiated from the partition directly;

W_{Ff}, which has entered flanking constructions and is radiated from flanking constructions; and

W_{leak}, which has been transmitted (as airborne sound) through leaks, ventilation ducts, and so on.

Also in this case, under the assumption of diffuse sound fields in the two rooms, the apparent sound reduction index may be evaluated from

$$R' = L_1 - L_2 + 10 \log \frac{S}{A} \quad \text{dB}.$$

The 10 log(S/A) correction term eliminates the receiving room reverberation effect and the effect of partition size, thus leading to the apparent sound reduction index that is not influenced by conditions surrounding the partition's use.

6.4 Traffic Noise Reduction Index

For exterior building facades and facade components, in situ noise reductions can be determined using traffic noise as the source.[34] The traffic sound reduction index R_{tr} is obtained from the equivalent sound pressure levels measured as a function of frequency on both sides of the test specimen (exterior facade or facade component). The traffic noise reduction index is denoted by R_{tr}:

$$R_{\mathrm{tr}} = L_{\mathrm{eq},1} - L_{\mathrm{eq},2} + 10\log\frac{S}{A}\quad \mathrm{dB},$$

where $L_{\mathrm{eq},1}$ is the equivalent sound pressure level 2 m in front of the test specimen, including the reflection effect of the test specimen, $L_{\mathrm{eq},2}$ is the equivalent sound pressure level in the receiving room averaged over the room, S is the area of the test specimen, and A is the equivalent absorption area in the receiving room in metric sabins. The above equation is applicable when the line of traffic is sufficiently long and straight to ensure a uniform distribution of incident sound. When the angle of elevation (observed from the point of least distance between the test specimen and the line of traffic) is more than about 20°, somewhat different results may occur because of oblique angles of incidence. (In this context, an angle of incidence of 0° is perpendicular to the facade. When the angle of elevation exceeds 50°, this equation should not be used.

6.5 Level Difference

The difference in the space and time average sound pressure levels produced in two rooms by one or more sound sources in one of them.[35] This quantity is denoted by D:

$$D = L_1 - L_2$$

where L_1 is the average sound pressure in the source room and L_2 is the average sound pressure level in the receiving room.

6.6 Standardized Level Difference

The adjusted level difference.[35] The adjustment is 10 times the logarithm of the ratio of the receiving room reverberation time and a reference reverberation time. This quantity is denoted by D_{nT}:

$$D_{nT} = D + 10\log\frac{T}{T_0}\quad \mathrm{dB},$$

where D is the level difference, T is the reverberation time in seconds in the receiving room, and T_0 is the reference reverberation time in seconds, which for dwellings is 0.5 s.

Normalizing (standardizing) of the level differences to a revelation time of 0.5 s takes into account the field experience that furnished rooms of dwellings have an average reverberation time of 0.5 s. It has also been observed that this reverberation time is nearly independent of room volume and frequency. This adjustment will result in the noise reduction, D_{nT}, for a pair of rooms of dissimilar volumes joined by a common wall to be different depending on whether the source and receiver rooms are the larger and smaller, respectively, or *vice versa*. The D_{nT} for the larger source room and small receiving room will be smaller than the D_{nT} for the smaller source room and larger receiving room.

Normalizing the level difference to a 0.5-s reverberation time in the receiving room is equivalent to normalizing the level difference to that with a receiving room having an equivalent absorption area of

$$A_0 = 0.32V,$$

where A_0 is the equivalent absorption area in square metres and V is the volume of the receiving room in cubic metres.

6.7 Standard Traffic Level Difference

Similarly, the in situ standardized traffic noise level difference $D_{nT,\mathrm{tr}}$ is defined as follows[36]:

$$D_{nT,\mathrm{tr}} = L_{\mathrm{eq},1} - L_{\mathrm{eq},2} + 10\log\frac{T}{T_0}\quad \mathrm{dB},$$

where T is the measured reverberation time in seconds in the receiving room and T_0 is the reference reverberation time, 0.5 s for dwellings.

6.8 Impact Sound Pressure Level

The average sound pressure level in a specific frequency band in the receiving room when the floor under test is excited by the standardized impact sound source.[37] This quantity is denoted by L_i.

6.9 Normalized Impact Sound Pressure Level

The impact sound pressure level L_i adjusted by a correction term in decibels equal to 10 times the common logarithm of the ratio between the measured equivalent absorption area A of the receiving room and the reference absorption area A_0.[37] This quantity is denoted by L_0:

$$L_0 = L_1 + 10\log\frac{A}{A_0}\quad \mathrm{dB},$$

where $A_0 = 10\ \mathrm{m}^2$.

In all cases, where it is uncertain whether results are obtained without flanking transmission, the normalized impact sound pressure level should be denoted by L_0'.

6.10 Standardized Impact Sound Pressure Level

The impact sound pressure level L_i adjusted by a correction term in decibels equal to 10 times the common logarithm of the ratio between the measured reverberation time T of the receiving room and the reference reverberation time T_0.[38] This quantity is denoted by L'_{nT}:

$$L'_{nT} = L_i - 10 \log \frac{T}{T_0}.$$

For dwellings T_0 is given by

$$T_0 = 0.5 \text{ s}.$$

Normalizing the impact sound pressure level to a reverberation time of 0.5 s accounts for the field experience that reverberation times in dwellings is nearly independent of the volume and frequency, and is equal to 0.5 s.

Normalizing of the impact sound pressure level to a reverberation time of 0.5 s is equivalent to normalizing the impact sound pressure level to an equivalent absorption area of

$$A_0 = 0.32V,$$

where A_0 is the equivalent area in square metres and V is the volume of the receiving room in cubic metres.

6.11 Reduction of Impact Sound Pressure Level (Improvement of Impact Sound Insulation)

The difference between the average sound pressure levels in the receiving room before and after installation of, for example, a floor covering; see ISO 140/VIII.[37] This quantity is denoted by ΔL.

6.12 Measurement and Evaluation of the Equivalent Absorption Area

The correction term for the above equation, containing the equivalent absorption area, may preferably be evaluated from the reverberation time measured according to ISO/R 354 and evaluated using Sabine's formula.[37]

$$A = \frac{0.163V}{T},$$

where A is the equivalent absorption area in square metres, V is the receiving room volume in cubic metres, and T is the reverberation time in seconds.

An alternative method of determining the equivalent absorption area is to measure the average sound pressure level produced by a known, sufficiently stable sound source.

6.13 Average Sound Pressure Level in a Room

Ten times the common logarithm of the ratio of the space and time average of the sound pressure squared to the square of the reference sound pressure.[35] The spatial average is over the entire room with the exception of those areas significantly influenced by direct radiation of a sound source or the near field of the boundaries (wall, etc.). This quantity is denoted by L:

$$L = 10 \log \frac{p_1^2 + p_2^2 + \cdots + p_n^2}{np_0^2} \quad \text{dB},$$

where $p_1, p_2, \ldots, p_n$ are the rms sound pressures at n different positions in the room and $p_0 = 20\ \mu\text{Pa}$ is the reference sound pressure.

REFERENCES

1. American National Standards Institute, "Rating Noise with Respect to Speech Interference," S3.14-1977 (R1986).
2. C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991, p. 17.9.
3. American National Standards Institute, "Procedure for the Computation of Loudness of Noise," S3.4-1980 (R1986).
4. "Procedures for Abatement of Highway Traffic Noise and Construction Noise," in Federal-Aid Program Manual, U.S. Department of Transportation, Federal Highway Administration, Washington, DC, August 9, 1982, Volume 7, Chapter 7, Section 3.
5. Massachusetts Department of Environmental Protection, Bureau of Waste Prevention–Air Quality, "Supplemental Form for Survey of Noise Potential," BWP AQ SFP 03.
6. City of Cambridge, Massachusetts Noise Ordinance, Ordinance No. 877, 1977.
7. "Protective Noise Levels—Condensed Version of EPA Levels Documents," U.S. Environmental Protection Agency, Publication No. EPA 550/9-79-100, November 1978.
8. "Noise Control and Compatibility Planning for Airports," U.S. Department of Transportation, Federal Aviation Administration, Advisory Circular AC 150/5020-1, August 3, 1983.

9. 24 CFR Part 51, U.S. Department of Housing and Urban Development, Environmental Criteria and Standards, 44 FR 40860, July 12, 1979; amended by 49 FR 880, January 6, 1984.
10. California Noise Insulation Standards, State Building Code (Part 2, Title 24, CCR), UBC Appendix Chapter 35, "Sound Transmission Control," December 1988.
11. Appendix Chapter 35, "Sound Transmission Control," Uniform Building Code, International Conference of Buildings Officials, 1988.
12. L. L. Beranek (Ed.), *Noise Reduction*, McGraw-Hill, New York, 1960, pp. 518–520.
13. T. J. Schultz, "Noise-Criterion Curves for Use with the USASI Preferred Frequencies," *J. Acoust. Soc.*, Vol. 43, No. 3, 1968, pp. 637–638.
14. *1995 ASHRAE Applications Handbook*, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, New York.
15. L. L. Beranek, "Application of NCB Noise Criterion Curves," *Noise Control Eng. J.*, Vol. 33, No. 2, 1989.
16. L. L. Beranek, "Balanced Noise-Criterion (NCB) Curves," *J. Acoust. Soc. Am.*, Vol. 86, No. 2, 1989.
17. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Test Method Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method," C 423-90a.
18. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Practices for Mounting Test Specimens during Sound Absorption Tests," E 795-91.
19. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Practice for Determining a Single-Number Rating of Airborne Sound Isolation for use in Multi-unit Building Specifications," E 90-90.
20. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Test Method for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions," E 90-90, p. 678.
21. American Society for Testing and Materials, *Section 4—1994 Annual Book of ASTM Standards*, ASTM E 413-87 (R1994), p. 715.
22. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Test Method for Measurement of Airborne Sound Insulation in Buildings, Annex A1. Measurement of Field Sound Transmission Loss," ASTM E 226-90.
23. G. C. Tocci, "A Comparison of STC and EWR for Rating Glazing Noise Reduction," *Sound Vib.*, October 1987, p. 32.
24. *Monsanto Acoustical Glazing Design Guide*, 3rd ed., 1995.
25. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Test Method for Laboratory Measurement of Impact Sound Transmission through Floor-Ceiling Assemblies Using the Tapping Machine," E 492-90, p. 721.
26. American Society for Testing and Materials, *Section 4—1991 Annual Book of ASTM Standards*, "Standard Test Method for Field Measurement of Tapping Machine Impact Sound Transmission through Floor-Ceiling Assemblies and Associated Support Structures," E492-90, p. 794.
27. International Standards Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements: Part 6—Laboratory Measurements of Impact Sound Insulation of Floors," Standard 140/6.
28. K. A. Hoover, *An Appreciation of Acoustics*, Berklee School of Music, Boston, MA.
29. American National Standards Institute, "Methods for the Calculation of the Articulation Index," ANSI S3.5-1969 (R1986).
30. T. Houtgast and H. J. M. Steeneken, "A Review of the MTF Concept in Room Acoustics and its Use for Estimating Speech Intelligibility in Auditor," *J. Acoust. Soc. Am.*, Vol. 77, No. 3, March 1985.
31. D. Davis and C. Davis, *Sound System Engineering*, 2nd ed., Howard W. Sams & Co., Indianapolis, IN, 1987, p. 235.
32. C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991, p. 16.16.
33. International Standardization Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements—Part III: Laboratory Measurements of Airborne Sound Insulation of Building Elements, ISO 140/III.
34. International Standardization Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements—Part V: Field Measurements of Airborne Sound Insulation of Facade Elements and Facades," ISO 140/V.
35. International Standardization Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements—Part IV: Field Measurements of Airborne Sound Insulation between Rooms," ISO 140/IV, 1978.
36. International Standardization Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements—Part V: Field Measurements of Airborne Sound Insulation of Facade Elements and Facades," ISO 140/V.
37. International Standardization Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements—Part VI: Laboratory Measurements of Impact Sound Insulation of Floors," ISO 140/VI.
38. International Standardization Organization, "Acoustics—Measurement of Sound Insulation in Buildings and of Building Elements—Part VII: Field Measurements of Impact Sound Insulation of Floors," ISO 140/VII.

95

SOUND PROPAGATION IN WORK SPACES

ULRICH J. KURZE

1 INTRODUCTION

In industrial halls and landscaped offices, sound distribution is often different from the sound fields in "Sabine" spaces with homogeneous and isotropic distribution of sound energy at every point in time and space.[1] When the ratio of room height to width or the ratio of width to length exceeds a value of about 3 or when most of the absorption is concentrated on one room surface (floor, ceiling, or wall), the reverberant sound field is no longer uniform in space but generally decays with increasing distance from the source. Knowledge about the spatial and temporal decay rate as a function of room shape, amount and distribution of absorption, and density of office equipment, machinery, or other fittings in the room is needed for the design and evaluation of noise control measures as well as for sound emission measurements on machinery in situ.[2]

Sound emanating from a source is reduced in intensity during propagation due to spherical spreading, air absorption, scattering or partial reflection at obstacles and diffraction at edges, or discontinuities of material surface properties. Indoors, many different paths must be considered for the sound traveling from the source to a particular receiver. Overlapping of rays at the receiver compensates for part of the spreading losses. Modern computers allow for quantifying the resulting difference between sound power level of the source and sound pressure level at the receiver by tracing thousands of rays or wavelets in auditoria and workrooms of any shape.[3] Uncertainties mostly follow from necessary, but not very safe, assumptions about specular or diffuse reflections and about the magnitude and angular dependence of the absorption at room boundaries. In noise control engineering, additional problems arise from the scattering of sound on the path between wall reflections and from the diffraction by noise barriers and other large obstacles. However, the overall shape of workspaces can often be approximated by one, two, or three pairs of parallel planes, thus simplifying the calculation procedures dealing with the room geometry.

An overview of the dependence of sound pressure level on distance in various rooms is given by Fig. 1. Close to the source, the effect of the room is small. At larger distances, the sound pressure level may decay at a rate up to 6 dB per doubling of distance if the line of sight to the receiver is not blocked and even more in rooms densely filled with machinery or in offices equipped with numerous screens. As shown in Fig. 1, a certain difference between sound pressure level L at the receiver and sound power level L_W of the source can be due either to the geometric spreading of the direct field, to the decay of a reverberant field between a rigid ceiling and absorptive scatterers, or to the absorption of an acoustic ceiling above a screen shielding the direct sound. There are enough uncertain or flexible parameters to adjust different calculation procedures to measured data.[4] Therefore, consistency with few measurements is generally not sufficient to support any complex computer program. Instead, some physical insight may help in designing quieter workplaces.

2 EMPTY ROOMS

2.1 Definitions

When sound is much more likely to hit a wall or the ceiling of a room than any other fitting inside of it, the

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

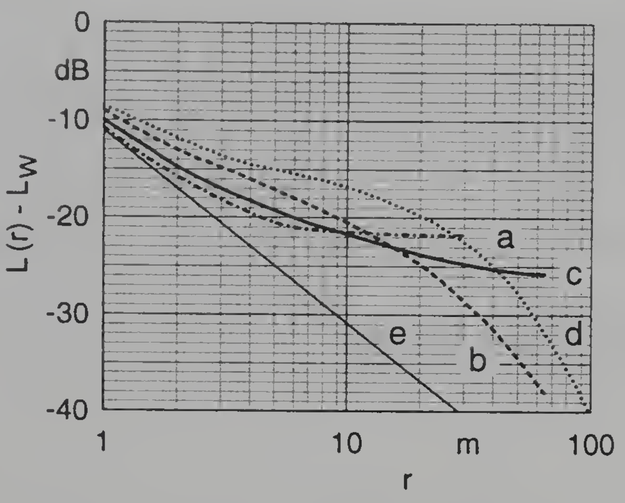


Fig. 1 Dependence of sound pressure level L minus sound power level L_W on distance r from a point source in different rooms with the same volume (27,000 m^3), little absorption ($\overline{\alpha} = 0.1$), and acoustically similar fittings ($l_S = 10$ m, $\alpha_{\text{fittings}} = 0.05$): (*a*) cubical room (30 × 30 × 30 m^3); (*b*) fitted flat room (73.5 × 73.5 × 5 m^3); (*c*) empty flat room; (*d*) fitted long room (270 × 10 × 10 m^3); (*e*) free field outdoors.

room may be considered empty. For quantification of this condition, one may relate the enveloping surfaces S of all fittings (machinery, office equipment, etc., without base area) in the room to the room's volume V and define the mean scattering density

$$q = \frac{S}{4V} \tag{1}$$

or the mean free path length

$$l_f = \frac{1}{q} \tag{2}$$

between two reflections. In empty rooms, this length is larger than twice the smallest dimension of the room (in many cases the room height). The concept can be applied to complete rooms or to uniform parts of differently fitted rooms. In principle, q and l_f depend on frequency because effective scattering is limited to wavelengths of sound that are smaller than the shortest perimeter of an obstacle. Since the A-weighted sound pressure level in offices and factories is very often determined by contributions in the octave bands centered at 0.5, 1, and 2 kHz, reasonable estimates for q or l_f follow, independent of frequency, from consideration of all obstacles with more than 1-m perimeter.

The absorption A of a room is defined by

$$A = \sum_i \alpha_i S_i, \tag{3}$$

where α_i is the Sabine absorption coefficient of the partial surface area S_i as obtained from standard reverberation room measurements.[5] In nondiffuse sound fields (empty flat and long rooms), the absorption coefficient of surfaces cannot be determined from reverberation time. Often estimated values must be used. For untreated walls, α lies in the range from 0.05 to 0.1 for engine rooms, 0.15 for general work shops, 0.2 for rooms in the wood-working industry, 0.2–0.25 for rooms of the textile industry, and 0.15–0.2 for offices.

In the frequency bands of interest the absorption coefficients listed for acoustic materials often exceed the value of 1. Such data cannot be used to determine reasonable reflection coefficients. The problem can be overcome by assuming that the listed data represent an average value of the true absorption coefficient α and the absorption exponent $\alpha' = -\ln(1-\alpha)$. The reflection coefficient ρ resulting from this assumption is shown in Fig. 2.

The absorption of equipment or machinery is rarely available from measurements. Absorption coefficients can be referred either to the enveloping surface or to the base area. For machinery with rigid surfaces it is often reasonable to assume an absorption coefficient $\alpha = 0.05$ for the enveloping surface or $\alpha = 0.1$ for the base area, independent of frequency.

The Sabine absorption coefficient describes an average material property for random sound incidence. If plane waves at a certain angle of incidence prevail, the angular dependence $\alpha(\vartheta)$ must be observed. For normal sound incidence the absorption coefficient $\alpha(0)$ is available from measurements in standing-wave tubes. For other directions, the assumption of a cosine law $\alpha(\vartheta) = \alpha(0)/\cos(\vartheta)$ is consistent with geometric acoustics and with constant temporal decay rates observed in Sabine spaces.[1]

2.2 Flat Rooms

The simplest case of the empty room is the flat room, where sound reflections from all side walls can be neglected. When floor and ceiling are plane parallel surfaces, the sound pressure level L in decibels re 20 μPa of the reverberant field can be calculated from consideration of specular reflections (see Fig. 3):

$$L = L_W + 10 \log \left(\frac{1}{2\pi r^2} \sum_{n_z=1}^{\infty} \cdot \frac{\exp[-r[m + \alpha'(0)/2h]\sqrt{1 + (2n_z h/r)^2}]}{1 + (2n_z h/r)^2} \right) \text{dB}, \tag{4}$$

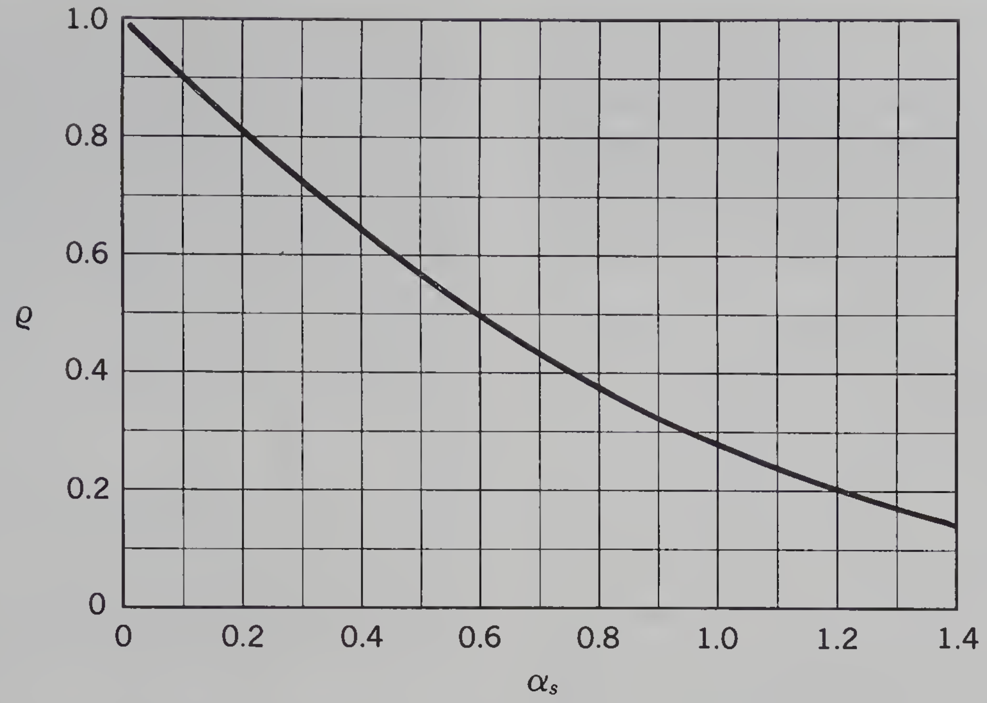


Fig. 2 Practical conversion from listed Sabine absorption coefficients α_S to reflection coefficients ρ required for calculation of spreading losses in reverberant fields, calculated from $\alpha_S = (1 - \rho - \ln \rho)/2$.

where L_W is the sound power level of a source on the floor in decibels re 1 pW, h is the room height in metres, m is the air absorption in nepers per metre or in decibels per 0.23 m, $\alpha'(0)$ is the absorption exponent of the ceiling for normal sound incidence, which is often similar to the Sabine absorption coefficient, and r is the horizontal distance from the center of the source in metres. Between a rigid floor and ceiling, the level L theoretically decays at a rate of 3 dB per doubling of distance r due to cylindrical spreading of sound energy as long as air absorption can be neglected. In reality, absorption coefficients between 0.05 and 0.1 must be assumed for most "rigid" surfaces (without acoustical treatment). Neglecting air absorption and the cosine law of absorption, the reverberant field can be approximated by the closed-form expression[6]

$$L \approx L_W + 10 \log \left(\frac{1}{4\pi r^2} \cdot \frac{5.7\rho_g}{[1.32(1-\rho_g) + h/r](1.93 + h/r)} \right) \quad \text{dB}, \quad (5)$$

where L_W is the sound power level of a source located halfway between floor and ceiling and ρ_g is the geometric mean reflection coefficient of floor and ceiling.

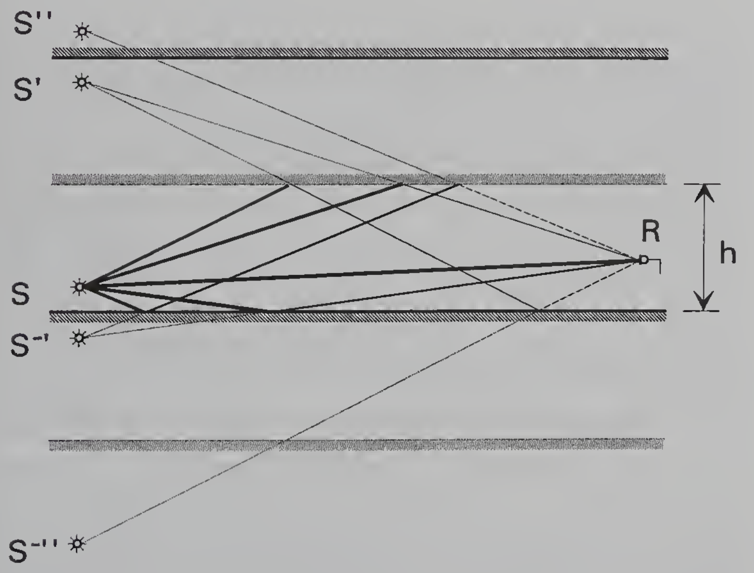


Fig. 3 Source and image sources in a flat room, ray constructions.

Reflections from inclined roofs, various ducts mounted at the ceiling, or beams and girders supporting the roof result in deviations from the laws of geometric acoustics. In such cases it is useful to apply Lambert's law for describing diffuse reflections from a rough surface (see Chapter 91). In optics this law states that the ray-bundle intensity of light decreases according to the cosine of the angle ϑ measured from the direction normal to the reflecting plane. Transformed to acoustics, it may be expected that the irradiation strength B_0 resulting from the sound incidence on a rough surface with area S and reflection coefficient ρ yields a power B_1 per unit area on the opposite surface, which follows from[7]

$$B_1 = \frac{\exp(-2mh)}{\pi} \int_S \frac{B_0 \rho \cos^2 \vartheta \, dS}{R^2}, \quad (6)$$

where R is the distance between the receiver on the surface 1 and the source element dS on the surface 0, $2mh$ is the air absorption in nepers over the mean free path length $2h$, and the reflection coefficient ρ is averaged over all angles of incidence. The solution of the integral equation following from Eq. (6) yields the approximate distribution of sound pressure levels of the reverberant sound field:

$$L \approx L_W + 10\log\left[\frac{1}{\pi h^2}\left(\frac{\rho_a}{(1+r^2/h^2)^{3/2}} + \frac{\rho_g^2}{1-\rho_g}\,\frac{b}{(b^2+r^2/h^2)^{3/2}}\right)\right] \text{ dB,} \tag{7}$$

where

$$b = 1 + \ln\left(1 + \frac{0.66}{1-\rho_g}\right). \tag{8}$$

Here, ρ_a is the arithmetic mean reflection coefficient of floor and ceiling and ρ_g is the geometric mean, respectively. In Fig. 4 results calculated with specular reflections from Eq. (5) are compared with those of Eq. (7). For diffuse reflections of the ceiling the sound pressure level close to the source is somewhat higher but it drops off faster at larger distances.

2.3 Long Rooms

Another case of interest is the empty corridor, the tunnel, or the duct with large cross section.[8] Geometric room acoustics employing image sources in a plane perpendicular to the duct axis yields attenuations with increasing distance by air absorption only as long as the walls are rigid. But as soon as some wall absorption or a reflection coefficient $\rho < 1$ is effective, air absorption plays a minor role and the sound pressure level at a distance x, which is large compared to the duct cross dimensions a and h, decays at the rate of 6 dB per doubling of distance, as in the free field of a point source:

$$L = L_W + 10\log\left(\frac{1}{4\pi}\cdot\sum_{n_y=-\infty}^{\infty}\sum_{n_z=-\infty}^{\infty}\frac{\rho^{|n_y|+|n_z|}}{x^2+(n_y a)^2+(n_z h)^2}\right) \text{ dB} \tag{9}$$

$$\approx L_W + 10\log\left[\frac{1}{4\pi x^2}\left(1+\frac{4\rho}{(1-\rho)^2}\right)\right] \text{ dB}$$

$$\text{for } x \gg a, h. \tag{10}$$

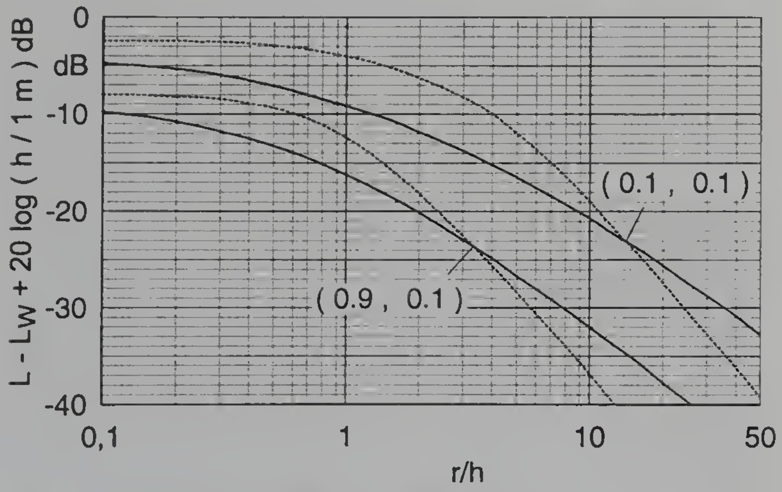


Fig. 4 Difference between sound pressure level L of the reverberant field and reduced sound power level $L_W - 20\log(h/1\text{ m})$ dB as a function of normalized distance r/h calculated for empty flat rooms of height h: (——) from Eq. (5) for specular reflection; (------) from Eq. (7) for diffuse reflection. Parameter: ceiling and floor absorption coefficients.

If the reflection coefficients of the four walls are different, ρ is taken as the area-weighted average,

$$\rho = \frac{a(\rho_2+\rho_4)+h(\rho_1+\rho_3)}{2(a+h)}, \tag{11}$$

where the walls reflecting with ρ_1 and ρ_3 are located opposite each other at a distance a and walls reflecting with ρ_2 and ρ_4 are normal to those two walls.

The sound propagation in a corridor with diffusely reflecting walls can be determined from the solution of the integral equation (see Chapter 91)

$$B(\mathbf{r}) = \frac{1}{\pi}\iint \rho(\mathbf{r}')B(\mathbf{r}')\,\frac{\cos\vartheta\cos\vartheta'}{R^2}\,dS' + B_0(\mathbf{r}), \tag{12}$$

where the vectors $\mathbf{r}$ and $\mathbf{r}'$ determine the position of two wall elements at a distance R, ϑ and ϑ' are the angles between rays and the direction perpendicular to the wall elements, and B_0 is the contribution of the direct sound. The integration has to be carried out over the entire wall surface. Results determined by numerical integration for a square duct are shown in Fig. 5. Similar to the curves for flat rooms, the reverberant sound pressure level is constant at very short distances and finally drops off at large distances, theoretically at a rate of 9 dB per distance doubling. It is important to realize that the total sound field composed of the direct and the reverberant contributions is dominated by the direct field both close to the source and at large distances. More examples for line sources across the duct, rectangular ducts with mixed specularly and diffusely reflecting surfaces, and round ducts can be found in the literature.[7,8]

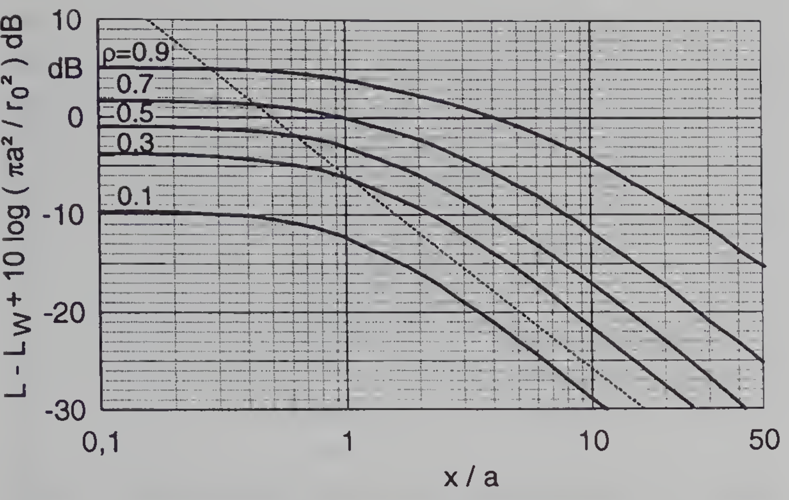


Fig. 5 Difference between sound pressure level L of the reverberant field and reduced sound power level $L_W - 10\log(\pi a^2/1\text{m}^2)$ dB as a function of normalized distance x/a calculated for empty long rooms of square cross section with area a^2 for various reflection coefficients ρ of the walls: (------) free field. (After Ref. 8.)

3 FITTED ROOMS

3.1 Low Density of Scatterers

When the room is filled with objects that are hit by sound rays traveling over the distance r with a small probability density[9,10]

$$w(r) = \frac{\exp(-r/l_f)}{l_f}, \tag{13}$$

one may combine the processes of scattering and wall reflection by the assumption that all objects are located directly at the walls. For rough walls the calculation procedure is the same as that for empty rooms except from a small increase of absorption due to the scatterers. For smooth plane walls, the total sound power must be split into two portions, one proportional to the probability of specular wall reflections $[q_W/(q_W+q_S)]$ and the other proportional to the probability of diffuse reflections $[q_S/(q_W+q_S)]$, where q denotes the reciprocal mean free path length, and the subscripts S and W refer to propagation paths between scatterers and walls, respectively. The propagation conditions for both portions can be calculated separately as described by Eq. (5), (7), or (10) for empty rooms before the contributions to the total sound field are finally added in terms of energies.

3.2 High Density of Scatterers

Since closed-form approximate solutions of the integral equation describing the diffuse sound field are available for certain isotropic spaces only and air absorption needs special consideration, alternative calculation procedures have been developed that are based on descriptions of diffuse sound fields in unlimited spaces. One such procedure employs the far-field approximation of the three-dimensional diffusion equation for weakly absorbing scatterers yielding the diffuse field sound pressure level[10]:

$$\lim_{r\to\infty} L \approx L_W + 10\log\left\{\frac{3q_s}{4\pi r} \cdot \exp\left[-q_s r\sqrt{3\left(\alpha_S' + \frac{m}{q_S}\right)}\right]\right\} \text{ dB.} \tag{14}$$

As compared to the divergence of wave fields with a decay rate of 6 dB per doubling of distance r, the diffuse field close to the source decays more slowly at a rate of 3 dB per doubling of distance and farther away from the source more rapidly due to the absorption α_S' of scatterers with density q_S and air absorption m.

Different procedures have been proposed to account for room boundaries with absorptive properties.[10–13] They are based on different concepts for the statistics of diffuse sound fields and for the appropriate averaging of absorption coefficients. Generally the results are not much different from each other and can all be found consistent with measured data depending on the assumptions of effective absorption coefficients and mean free path lengths. Widely used is the calculation method developed by Jovicic,[11] which only extends the exponential factor of Eq. (14) in order to describe the diffuse sound field in a flat room:

$$L = L_W + 10\log\left\{\frac{3q_S}{4\pi r} \cdot \exp\left[-q_S r\sqrt{3\left(\alpha_S' + \frac{m}{q_S} + b'\right)}\right]\right\} \text{ dB,} \tag{15}$$

where $b' = b'(\alpha_{\text{floor}}) + b'(\alpha_{\text{ceiling}})$ and

$$b'(\alpha) = -\ln\left(qh\left\{\left(1 - \frac{\alpha}{4}\right) + \left(1 - \frac{\alpha}{2}\right)\frac{2}{\alpha'} \cdot \left[1 - \exp\left(-\frac{\alpha'}{2}\left(\frac{1}{qh} - 1\right)\right)\right]\right\}\right) \quad \text{for} \quad qh < 1,$$

$$= -\ln\left(1 - \frac{\alpha}{4qh}\right) \quad \text{for} \quad qh \geq 1.$$

As far as needed, reflections from side walls must be taken into account by consideration of reflection coefficients ρ_{wall} and image sources in the horizontal plane. Furthermore, the complete description of the sound field requires the superposition of the direct field

$$L_{\text{direct}} = L_W + 10\log\left(\frac{1}{4\pi r^2}\exp[-(q+m)r]\right) \quad \text{dB} \tag{16}$$

from the original source and all the relevant image sources.

4 NOISE CONTROL MEASURES

4.1 Absorbers

The described calculation procedures allow one to predict the reduction of sound pressure level in rooms due to absorptive linings of walls and ceiling or sound-absorbing baffles. Considering the sound from neighboring work places exclusively, which is often a useful concept for the assessment of the acoustical quality of workspaces, one can obtain level reductions on the order of 5 dB. The effect is stronger in fitted than in empty rooms. Close to noisy work places, the effect is smaller due to the unchanged direct sound field.

The formulas give no indication for optimum configuration and location of absorptive materials. Obviously, image sources corresponding to wall reflections are most effectively attenuated by absorbers positioned close to the source. But noise sources are often distributed over larger areas. Consequently, modern factory halls are fitted with absorptive panels at the entire ceiling. In model studies, flat linings of the ceiling and suspended baffles showed no difference in attenuation performance. Other field measurements indicate some improvement due to low-hung baffles that reduce the free line of sight from the source to remote receivers. In all cases the positioning of the absorbers can be adjusted to the requirements for illumination, ventilation, sprinklers, and so on, without severe acoustical effects.

4.2 Screens

Absorptive noise screens provide for shielding from sound rays that otherwise reach the receiver on a straight path, for reduction of mean free path length (thus enhancing the noise reduction performance of absorptive ceilings), and for absorption of sound close to noise sources.[14] Neglecting the absorptive properties of the screen, one finds the minimum performance of a screen in terms of the insertion loss D_e as

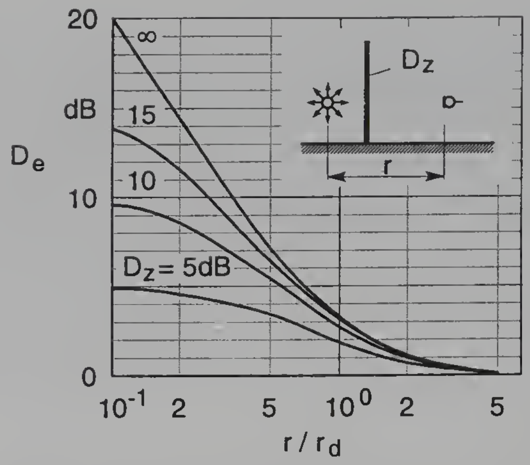


Fig. 6 Insertion loss D_e of a barrier in a reverberant field as a function of normalized distance calculated from Eq. (17). Parameter: barrier attenuation D_z.

$$D_e = 10\log\left(\frac{1 + r^2/r_d^2}{10^{0.1D_z} + r^2/r_d^2}\right) \quad \text{dB}, \tag{17}$$

where r_d is the distance at which the energy densities of the direct and the reverberant sound are equal and D_z is the attenuation by diffraction, which can be calculated from the path length difference z between diffracted and direct rays relative to the wavelength λ of sound. Indoors, one may conservatively assume a relation $D_z = 10\log(1 + 5z/\lambda)$ dB instead of the 6-dB higher values in free space. A graph showing the decreasing attenuation performance of screens with increasing distance is given in Fig. 6.

Results of numerous field measurements in flat halls are summarized in Table 1. Close to high screens, the insertion loss was found as high as 10 dB. But at some distance and for relatively low screens, the values drop below 4 dB.

5 MEASUREMENT PROCEDURES

5.1 Acoustical Quality of Workspaces

In Sabine spaces the acoustical quality can be expressed in terms of the reverberation time. In flat and long rooms,

TABLE 1 Insertion Loss D_e, in Decibels, of Barriers in Factory Halls, Mean Values, and Standard Deviations of Data Measured in Octave Band Centered at 1000 Hz

	Distance from Source to Receiver/Room Height		
Barrier Height/ Room Height	0.3	0.3–1	1–3
<0.3	7.4 ± 1.4	3.6 ± 2.1	—
0.3–0.5	10	7.1 ± 1.8	4.5 ± 1.8
>0.5	—	8.6 ± 1.7	6.3 ± 1.5

the reverberation time depends on the distance between source and receiver and strongly on the mean free path length between scatterers. The absorption coefficient determined from reverberation time measurements in such rooms is generally smaller than data reported from laboratories and reaches such data in densely fitted rooms only. The performance of acoustical treatment of room surfaces is better evaluated from the attenuation obtained at certain distances from the sound source, for example, 5, 10 and 20 m, in the frequency bands of interest, for example, 500, 1000, and 2000 Hz.[15] Preferably, a reference sound source with uniform directivity and known sound power level is used for the measurements before and after acoustical treatment. If no measurements have been carried out before, the difference between sound power level and sound pressure level can be evaluated and compared to theoretical data. The evaluation of level gradients, for example, level differences per distance doubling, which can be done simply without knowledge of the sound power level, is generally not useful because the gradients may strongly vary with distance and depend more on the geometry of the room and its fittings than on absorption.

To determine the minimum performance of acoustic ceiling or wall treatments, measurements are preferably carried out with the sound source and the microphone on a free passageway not far from a wall. The performance of noise barriers should be measured at the work places actually affected. As to the selection of measurement positions, some guidance may be obtained from the International Standardization Organization (ISO) standard 10053.[16] However, the insertion loss derived from this standard refers to free-field conditions that provide a much higher attenuation performance.

5.2 Sound Emission of Machinery

Two types of data are used to describe the noise emission by machinery and equipment. One is the sound power level (preferably in frequency bands) according to ISO 3740 or ISO 9614 and the other is the A-weighted sound pressure level at the operator's ear according to ISO 11200. Both emission data are referred to free-field conditions. In situ measurements must be corrected for room effects. Best results are obtained when the room effects are kept small. This can be achieved by selecting measurement positions close to the source (either microphones on an enveloping surface at small distances of 0.5–1 m from the machine surface or vibration pick-ups at the surface and estimated radiation efficiencies), directive microphones (which are difficult to calibrate in situ), or sound intensity techniques. Alternatively, the room effects may be determined from measurements with a reference sound source of known sound emission that replaces the unknown source.

REFERENCES

1. L. Cremer, H. A. Müller, and T. J. Schultz (Transl.), *Principles and Applications of Room Acoustics*, Applied Science, Vol. 1, London, New York, 1982.
2. International Standardization Organization, "Acoustics—Determination of Sound Power Levels of Noise Sources—Engineering Method for Free-Field Conditions over a Reflecting Plane," ISO 3744, 1981.
3. A. M. Ondet and J. L. Barbry, "Modeling of Sound Propagation in Fitted Workshops Using Ray Tracing," *J. Acoust. Soc. Am.*, Vol. 85, No. 2, 1989, pp. 787–796.
4. H. Lazarus, "Schallausbreitung in Flachräumen" ("Sound Propagation in Flat Rooms"), Fortschritte der Akustik—DAGA '87, Bad Honnef: DPG-GmbH 1987, pp. 157–178 (in German).
5. International Standardization Organization, "Acoustics—Measurement of Sound Absorption in a Reverberant room," ISO 354, 1985.
6. H. Schmidt, *Schalltechnisches Taschenbuch* (*Handbook of Acoustics*), 4th ed., VDI-Verlag, Düsseldorf, 1989 (in German).
7. H. Kuttruff, "Stationäre Schallausbreitung in Flachräumen" ("Stationary Sound Propagation in Flat Rooms"), *Acustica*, Vol. 57, 1985, pp. 63–70 (in German).
8. H. Kuttruff, "Stationäre Schallausbreitung in Langräumen" ("Stationary Sound Propagation in Long Rooms"), *Acustica*, Vol. 69, 1989, pp. 53–62 (in German).
9. H. Kuttruff, *Room Acoustics*, 2nd ed., Applied Science, London, 1979, p. 102.
10. U. J. Kurze, "Scattering of Sound in Industrial Spaces," *J. Sound Vib.*, Vol. 98, No. 3, 1985, pp. 349–364.
11. German Standardization Organisation, "Computation and Measurement of Sound Propagation in Workrooms" (in German), VDI 3760. 1993.
12. E. A. Lindqvist, "Noise Attenuation in Factories," *Appl. Acoust.*, Vol. 16, 1983, pp. 183–214.
13. M. Hodgson, "Theoretical and Physical Models as Tools for the Study of Factory Sound Fields," Ph.D. Dissertation, Southampton, 1983.
14. U. J. Kurze, "The Performance of Noise Barriers in Open-Plan Offices and Industrial Buildings," *Noise Control Eng.*, Vol. 11, No. 3, 1978, pp. 116–123.
15. International Standardization Organization, "Acoustics—Guidelines for the Design of Low-Noise Work Places," 1992.
16. International Standardization Organization, "Acoustics—Measurements of Office Screen Sound Attenuation under Specific Laboratory Conditions," ISO 10053, 1991.

96

ACOUSTICAL GUIDELINES FOR BUILDING DESIGN

EWART A. WETHERILL

1 INTRODUCTION

This chapter summarizes the basic requirements of building acoustics and enumerates suitable construction and interior finish materials for the situations most likely to be encountered. It provides a working guide to enable a designer or builder with little or no background in acoustics—but with the patience to follow a sequence of well-defined steps—to distinguish between problems of *noise intrusion* and *room acoustics*, to evaluate the potential for an acoustical problem, to consider alternative solutions, and to recognize when more rigorous evaluation is needed.

Successful acoustical design of buildings requires familiarity with appropriate construction methods, and also with building codes and other local restrictions, to ensure full compatibility before construction begins. When a conflict in priorities becomes apparent during construction, the builder will seldom place hard-to-prove acoustical needs ahead of better understood and more tangible construction needs.

Most acoustical questions can be resolved by following the steps illustrated in Fig. 1 for which the designer selects appropriate criteria for the particular building or space to determine the required type of construction. This review procedure also defines plan relationships that should be avoided and provides a checklist for coordinating the work of other disciplines and examining the appropriateness of possible design revisions. Detailed references are found in other chapters as noted herein.

2 ACOUSTICAL DESIGN OF BUILDINGS

Architectural acoustics encompasses the process through which acoustical conflicts can be *avoided*, unwanted sounds *controlled*, and wanted sounds *enhanced* by the building design. Any building may contain many spaces, each with its own requirements for acoustical quality and for isolation from noise and vibration.

The acoustical design for a particular building, or for a space within the building, must be compatible with other construction requirements and must be incorporated in the architectural drawings and specifications that form the basis for the awarding of a construction contract. This process requires close coordination and detailed cross checking because virtually any architectural design decision can have acoustical implications. The design of a building is generally a series of compromises in which benefits for one discipline must be weighed against disadvantages for another. The designer should know to what extent acoustical goals may be compromised without significantly affecting the usefulness of the building. Basic design assumptions should be firmly established before consideration of detailed requirements. For example, how important is total control of exterior noise or speech privacy between spaces? Or, should the interior be designed to *enhance* or *suppress* projection of sound? Clarification of such design issues at the outset will simplify the detailed selection process.

Section 6 summarizes general concerns and acoustical design requirements for current building types in several categories. If the proposed occupancy encompasses two or more categories, each one should be considered and the more stringent acoustical requirements should govern the design.

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

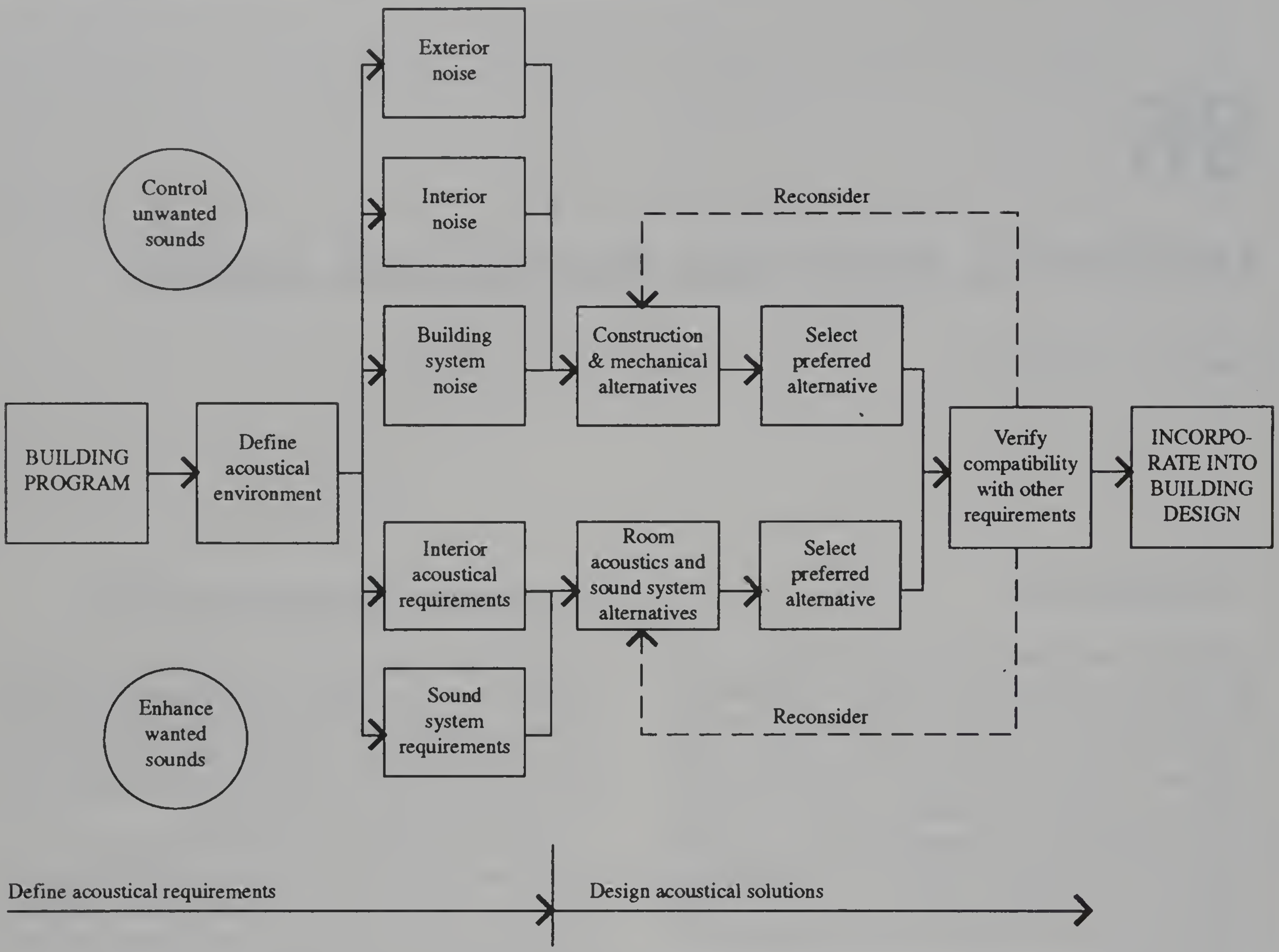

Exterior noise
Interior noise
Building system noise
Control unwanted sounds
BUILDING PROGRAM
Define acoustical environment
Construction & mechanical alternatives
Select preferred alternative
Reconsider
Verify compatibility with other requirements
INCORPO-RATE INTO BUILDING DESIGN
Interior acoustical requirements
Sound system requirements
Room acoustics and sound system alternatives
Select preferred alternative
Reconsider
Enhance wanted sounds
Define acoustical requirements
Design acoustical solutions

Fig. 1 Diagram illustrating steps to be used to solve acoustical questions.

3 CRITERIA

3.1 Selection of Acoustical Criteria

The acoustical criteria for a facility—or for individual spaces within it—can be stated *objectively* in terms of physical properties required to meet a specified need. However, they may be conditioned by a wide range of *subjective* requirements such as a listener's hearing acuity or individual sensitivity to intruding noise, as well as by cost and other nonacoustical requirements. Basic acoustical concerns are background noise level, reverberation, and noise levels generated within the listening space (see Chapter 94).

The desired background noise level for each space is typically specified by reference to noise criteria or other rating systems (see Chapter 80). Selection of desired background level is determined by the following:

(a) Required degree of communication between users

(b) Speech privacy required from adjacent spaces or nearby occupants of the same space (e.g., open-plan offices)

Actual background level is governed by the following:

(a) Control of intruding noises (from outdoors or from adjacent spaces) by appropriate construction (see Chapter 93)

(b) Control of noise from building systems (including air conditioning, plumbing, elevators, etc.) (see Chapter 98)

(c) Use of sound masking system to control minimum sound level (see Chapter 163)

Reverberation time (see Chapters 92 and 94) is an adequate single descriptor for the acoustical quality of most spaces. It is governed by the following:

(a) Volume and shape of space

(b) Amount and location of sound absorption in space

A quick estimate of the reverberation time inherent in a space of a given size will enable the designer to approximate the amount of sound absorption that is needed. Section 5.1 shows how to calculate approximate reverberation time, while Table 4 defines recommended mid-frequency (i.e., 500–1000 Hz) reverberation times for most common uses. Detailed calculations of reverberation time may be required in the final design and will be required during preliminary design for spaces where acoustical quality is of particular importance.

For acoustically sensitive uses such as music performance and for other special requirements, reverberation time alone may not be an adequate criterion; thus a more detailed analysis is needed. Spaces in which *projection* of sound is required, such as auditoria, must be treated differently from spaces in which *nonprojection* of sound is required (e.g., library reading room) (see Chapters 95 and 163).

3.2 Anticipated Noise Levels within the Space

In addition to the need for freedom from intruding noises, the likelihood that loud sounds could be generated within the space should be considered. For example, in a band rehearsal room special sound-absorbing treatment will be required to control the sound of the band (see Chapter 92).

3.3 Selection of Acoustical Requirements

Sections 4–6 enable the designer to select the acoustical conditions appropriate for a specific use, to assess the intruding noise or vibration that must be controlled, and to define the interior acoustical requirements of the space as summarized in Fig. 1. Using the work sheet in Section 7 and illustrative examples in Section 8, the designer selects the type of construction and interior treatment required to meet the design goals. To accommodate nonacoustical criteria, the designer applies a judgment factor, based on familiarity with the requirements for the particular building or space, to choose between possible alternatives.

Recommended values for background noise and reverberation are presented, in addition to subjective criteria, which are summarized for typical occupancies in Section 6. Each step in the evaluation process requires the designer to select the most appropriate condition from a range of five alternatives on the basis of familiarity with the project requirements and site conditions. These alternatives will be readily understood by most designers or can be approximated by reference to existing buildings. Where extreme conditions are encountered, or if considerable doubt exists as to the appropriate selection, the designer is advised to seek the assistance of an experienced consultant.

While this summary is based on generally accepted practice in acoustical design and on current North American building codes, the designer should also check on local requirements and on special conditions such as defined by the Americans with Disabilities Act (ADA), Occupational Safety and Health Act (OSHA), and so forth (see Chapters 119 and 120).

4 SELECTION OF CONDITIONS FOR SPACE BEING ANALYZED

This section presents criteria for noise and vibration (Table 1) and sources of intruding noise and vibration (Table 2).

5 CONSTRUCTION ALTERNATIVES FOR SPACE BEING ANALYZED

This section presents criteria for the control of intruding noise and vibration (Table 3).

5.1 Estimation of Reverberation Time

Approximate reverberation time (RT) in seconds can be derived using the Sabine equation (for detailed analysis refer to Chapters 77 and 92:

$$\mathrm{RT} = \frac{KV}{A_1 + A_2 + A_3 + \cdots},$$

where K is 0.16 (SI units) or 0.05 (English units), V is the volume in cubic metres or cubic feet, and A is the absorption coefficient α of each material multiplied by its surface area. For a preliminary estimate, absorption coefficients can be simplified:

For hard materials (e.g., concrete, wood) assume $\alpha = 0.1$.

For soft materials (thick carpet, "acoustical" materials, etc.) $\alpha = 0.8$.

As an example consider an auditorium with a volume of 807.5 m^3. Also assume the following:

Area 1			
Ceiling, walls, exposed floor	417.8 m^2	$\alpha_1 = 0.1$	$A_1 = 41.78$
Area 2			
Empty seating (not upholst.)	133.8 m^2	$\alpha_2 = 0.1$	$A_2 = 13.38$
Area 3			
Sound-absorbing material	25.6 m^2	$\alpha_3 = 0.8$	$A_3 = 20.48$
		Total empty	75.64

TABLE 1 Criteria for Noise and Vibration Based on a Range of Five Alternatives

1	2	3	4	5
a. Sensitivity to noise: desired background noise levels (Chapter 94)				
Not critical, e.g., shopping mall (NC 45–50)	Restaurant, open office (NC 35–45)	Office, quiet recreation (NC 25–35)	Sleep, worship, performance (NC 20–30)	Special uses: recording, test laboratory (<NC 15–20)
b. Sensitivity to vibration (Chapters 85 and 93)				
Not critical	Possible distraction	Sensitive to traffic vibration, footfalls	Laboratories/ optical processes	Extremely vibration sensitive
c. Speech privacy requirements (Chapter 94)				
No concern	Conversational	Residential	Confidential	Security
d. Anticipated sound level within space (Chapters 86 and 94)				
Conversation, general activity	Teaching, laboratory, computer use	Quiet recreation, music performance and playback	Indoor sports, mechanical equipment, industrial	Rock music, very high sound levels
e. Anticipated vibration within space (Chapters 85 and 93)				
Footfalls	Plumbing, home appliances	Mechanical equipment and piping	Aerobics, dancing	Industrial processes, heavy machinery

TABLE 2 Sources of Intruding Noise and Vibration Based on a Range of Five Alternatives

1	2	3	4	5
a. Exterior noise and vibration sources (Chapters 87 and 88)				
Rural site, extremely quiet	Suburban, residential	Suburban, near major street, outdoor equipment	Urban site near heavy-traffic streets	Urban, truck traffic, industry, rail, airport
b. Interior sound sources: adjacent space (Chapters 86 and 94)				
Conversation, general activity	Teaching, laboratory, computer	Quiet recreation, music performance and playback	Indoor sports, mechanical equipment, industrial	Rock music, very high sound levels
c. Interior vibration sources (Chapters 85 and 93)				
Footfalls	Plumbing, home appliances appliances	Mechanical equipment and piping	Aerobics, dancing	Industrial processes, heavy machinery
d. Noise and vibration from HVAC[a] systems (Chapter 98)				
Remote central system, low velocity air	Central equipment room, medium- to high-velocity air system	Equipment suspended above ceiling: fan-coil, air terminal boxes	Adjacent mechanical room, roof-mounted equipment above	Special-use systems, high noise and vibration
e. Other building equipment (Chapters 85 and 93)				
Plumbing, domestic water systems	Pneumatic tube systems, automatic door openers	Elevators (especially hydraulic), escalators	Adjacent mechanical equipment room	Special equipment, e.g., fountain pump, emergency generator

[a] Heating, ventilation, and air conditioning.

TABLE 3 Control of Intruding Noise and Vibration Based on a Range of Five Alternatives

1	2	3	4	5
a. Exterior walls				
No acoustical concerns, open windows	Lightweight construction, windows closed	Stucco or heavier, laminated or double glazing	Masonry walls, reduced window area	Double-wall construction, no windows
b. Interior walls				
No acoustical concerns, studs and gypsum board	Lighweight construction, no openings	Meet code requirements for multiunit dwellings	Composite wall, preferably masonry	Double-wall construction, (consider relocation)
c. Roof/ceiling				
No acoustical concerns, ceiling optional	Lightweight construction, acoustic tile or gypsum board suspended ceiling	Lightweight concrete deck with suspended gypsum board ceiling	Concrete deck with resiliently suspended ceiling, minimum 36-in. cavity	Heavy concrete deck with fully isolated heavy ceiling
d. Floor/ceiling				
No acoustical concerns, ceiling optional	Lightweight construction, acoustic tile or gypsum board suspended ceiling	Meet code requirements for multiunit dwellings	Concrete deck with resiliently suspended heavy ceiling	Double (floated) floor with resiliently suspended heavy ceiling consider relocation)
e. Control of vibration/impacts				
No acoustical concerns	Lightweight construction, with suspended ceilings	Meet code requirements for multiunit dwellings	Vibration sources resiliently isolated	Complete isolation from structure with vibration source
f. Control of HVAC noise/vibration (Chapter 98)				
Quiet fan selection, resilient mounting	Duct liner, low air velocity, resiliently isolated equipment	Equipment enclosed, resilient isolators, main ducts remote	Special construction and resilient isolators, ducts lagged or enclosed	Equipment completely remote from occupied spaces
g. Control of noise from building systems (Chapters 85 and 86)				
Resilient sleeving at all piping connections	Remote equipment, enclosed and resiliently isolated	Elevators, trash chutes remote, equipment/noise and vibration isolated	Heavy construction, equipment fully isolated	Equipment completely remote from occupied spaces

Note: See Chapter 93.

Add a full audience:

$$133.8\ \text{m}^2 \qquad \alpha = 0.7(0.8 - 0.1) \qquad \text{Add } \underline{93.66}$$

$$\text{Total full} \quad 169.30$$

Then

$$\text{RT empty} = \frac{0.16 \times 807.5}{75.64} = 1.71 \text{ s.}$$

$$\text{RT full} \ \frac{0.16 \times 807.5}{169.30} = 0.76 \text{ s.}$$

Table 4 presents criteria for the control of sound within a space.

TABLE 4 Control of Sound within the Space Based on a Range of Five Alternatives

1	2	3	4	5
		a. Spaces where nonprojection of sound required (Chapter 92)		
Dwelling; RT 0.5–0.7 s	Work, study; RT 0.8–1.0 s	Dining, assembly, multiuse spaces; RT 0.8–1.0 s	Industrial processes, high noise levels; RT 0.8–1.0 s	Special low noise, recording, anechoic; RT 0.4–0.6 s
		b. Spaces where projection of sound required (Chapter 95)		
Drama, high speech clarity; RT 0.4–0.8 s	Conference, lecture, classroom; RT 0.6–0.8 s	Multipurpose audience, musicals; RT 1.2–1.6 s	Opera, chamber orchestra, ballet; RT 1.4–1.7 s	Orchestra, organ, chorus; RT 1.7–2.2 s

5.2 Shaping of Space for Optimum Sound Distribution

This important aspect of auditorium design requires much greater detail than can be provided in this handbook. The reader is referred to the references listed at the end of Chapter 90 for further information.

6 BUILDING CATEGORIES

6.1 Dwelling

The most important acoustical requirement in dwellings is control of intruding noise, particularly at night, when the ambient sound level is lowest. Private homes typically rely upon open windows for ventilation and so are sensitive to outdoor noises. Reverberation is generally controlled by household furnishings, and other internal acoustical conditions are seldom of concern. (See Table 5.)

For new dwellings in noisy urban settings or close to heavy traffic, airports, and so forth, some jurisdictions require special construction for noise control, which also entails mechanical ventilation so that occupants can leave windows closed to limit noise intrusion. These are typical requirements for apartment/condominium construction, in addition to control of airborne and structure-borne noise from adjacent dwelling units and from building systems (e.g., plumbing, heating, trash chutes). However, noise control standards in current building codes do not ensure satisfactory noise and impact isolation between dwelling units. Recent studies suggest that noise and impact isolation required to avoid disturbance of neighbors cannot be satisfied by light frame construction.

Hotel noise control requirements are frequently exacerbated by proximity to airports or major highways. Internally, adequacy of noise isolation between guest rooms, building systems noise control, and isolation of guest rooms from meeting and convention facilities must be considered.

Hospital wards require stringent noise control measures because bed-ridden patients cannot escape from intruding noises. Special attention should be given to

TABLE 5 Acoustical Requirements for Dwelling Spaces

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time		
Required Use of Space	Desired Background Level	Sensitivity to Vibration	Speech Privacy Required	Maximum Sound Level in Space	Maximum Vibration in Space	Nonprojection of Sound	Projection of Sound Required	Is Sound Amplification Required?
Private Residence	4	1	3	1	2	1		
Apartment/Condo	4	1	3	1	2	1		
Hotel guest rooms	4	1	3	1	3	1		
Hospital Ward	4	2	3	1	3	1		

[a] See Table 1.

[b] See Table 4.

planning and design for control of noise and vibration from hospital activities, building systems, shared bathrooms, and so forth. Despite the need for washable wall and floor surfaces, a substantial amount of sound-absorbing treatment is needed in corridors, nurses' stations, and public spaces to allow speech privacy and to minimize disturbance of patients.

6.2 Work/Study

Acoustical requirements for work or study facilities depend on the tasks to be accommodated (Table 6). For open-plan offices, the background level should be low enough for telephone use yet high enough for some degree of speech privacy between workstations. Since the background level from variable air volume (VAV) ventilation systems can change significantly as air quantity changes, it may be necessary to add an electronic background sound system to maintain a constant sound level to enhance speech privacy. Floor and ceiling finishes should be sound absorbing to minimize sound projection, while acoustical separation between workstations should be provided by sound-absorbing partial-height partitions. Enclosed offices may require a higher degree of privacy.

Large open-plan spaces such as library reading rooms and museum galleries require nonprojection of sound, that is, localization of activity sounds and control of reverberation. In general, a sound-absorbing ceiling is a basic requirement, with a carpeted floor or wall-mounted sound-absorbing treatment and with a constant background sound level for effective masking of intruding noises. Quiet areas for serious study and meditation should be as remote as possible from circulation spaces, reference desks, and other noisy activity areas.

Mechanical equipment for air conditioning and ventilation should be well removed from occupied spaces wherever possible. However, for cost-saving reasons, roof-mounted air conditioning units and noisy VAV boxes may be planned directly above acoustically sensitive spaces. In such situations, noise and vibration could be virtually impossible to control. Main ducts close to large fans should be kept away from occupied spaces for control of "breakout" noise through duct walls. As a general rule, roof-mounted units and VAV boxes should be located above storage or washrooms, with acoustical lining in ducts serving occupied spaces.

6.3 Meeting/Hospitality

Meeting spaces may vary from fixed-seating, special-purpose facilities such as lecture and demonstration rooms to flat-floor, divisible rooms that are easily adapted to many types of events (Table 7). For either extreme, the size may vary from a classroom for 25–30 people to a convention facility seating over a thousand people, with amplified sound and projection systems. In general, sound amplification is desirable for all except the very smallest spaces, while standards such as ADA make hearing-impaired systems a special requirement for many uses. Where video demonstrations and sound playback are required, projection and lighting systems should be integrated with sound amplification. Microphones may be required in the seating area for audience

TABLE 6 Acoustical Requirements for Work/Study Spaces

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Office								
Open plan	2	1	1	2	1	2	—	—[c]
Closed plan	3	1	2	1	1	2		
In noisy building	3	1	1	2	3	2	—	—[c]
High security	3	1	5	1	1	2		
Library								
Reading	3	2	2	1	1	2		
Reference	3	1	2	1	1	2		
Classroom	3	1	2	1	1	—	2	
Museum	3	1	1	1	1	2		
Meditation	2	2	4	1	1	1		

[a] See Table 1.

[b] See Table 4.

[c] Special evaluation needed.

TABLE 7 Acoustical Requirements for Meeting/Hospitality Spaces

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Lecture/demonstration rooms	4	1	3	3	1	—	2	Y
Conference	4	1	4	3	1	—	2	Y
Dining	2	1	2	1	1	3	—	—
Assembly	3	2	2	3	1	3	—	Y
Divisible space	3	2	3	3	1	3	—	Y

[a] See Table 1.
[b] See Table 4.

response. Fixed lecture/demonstration spaces could also enhance unamplified speech or music if sound reflecting wall and ceiling surfaces are suitably shaped for optimum natural projection of reflected sounds. Reverberation times should be low for speech and amplified sound but could be lengthened for music performance by use of retractable curtains or other removable sound-absorbing materials.

By contrast, multiuse spaces such as divisible convention rooms should be designed for nonprojection of unamplified sound, should have substantial areas of fixed sound absorption on ceiling and upper walls for control of crowd noise, and should rely upon sound amplification for most events. It is practically impossible to design for natural sound reinforcement from walls and ceiling if the room size and sound source location may vary and if large movable wall surfaces must be flat and acoustically untreated. The sound system layout should be coordinated with the divisible sections of the facility so that each subspace can function either independently or as part of the larger space.

6.4 Performance

The most difficult of acoustical challenges, performing arts facilities range from low-cost flexible spaces such as community centers or drama/music classrooms to highly specialized auditoria seating 3000 or more (Table 8). These may be single-purpose spaces or may have complex stages that can be changed mechanically to accommodate from orchestra and chorus to opera/ballet or drama. Each type of event has well-defined requirements for sight-lines and lighting. Reverberation requirements may vary from under 1.0 s for drama to over 2.0 s for choral/orchestral and organ works. Very few existing major facilities provide truly optimum conditions for the full range of performing arts, mostly because of the high cost of large-scale adjustability. Detailed information on

TABLE 8 Acoustical Requirements for Performance Spaces

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time[c]		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Community center	4	2	2	5	4	3	2	Y
Theatre/opera	5	3	4	5	4		1/4	—[c]
Multipurpose auditorium	4	2	4	5	4		3	Y
Orchestral	5	3	4	5	2		5	—[c]
Rehearsal	4	2	4	5	4		1/2	
Outdoor performance	4	1	1	5	4			Y

[a] See Table 1.
[b] See Table 4.
[c] Special evaluation needed.

design of performing arts facilities can be found in the references in Chapter 91.

The primary acoustical requirement for a performance space is a low background noise level, free from distracting noise intrusions, to allow the audience to hear stage whispers and pianissimos. This usually determines the construction of the building enclosure and may also lead to rejection of a building site that is too noisy. The ventilation system must be conservatively designed, with remote mechanical equipment spaces and very low, constant air velocity so that diffusers are completely silent. These requirements must be preserved even when cost cutting is necessary.

After performance and audience areas are established (including seat spacing, aisle widths, and other code-mandated requirements), plan and section configurations should be reviewed to ensure the best compromise between sight lines, lighting, and acoustics. The total volume of the space (e.g., ceiling height) must allow for the required maximum reverberation time, avoiding deep underbalcony seating areas that do not share the acoustical characteristics of the main volume. In general, the best contribution to cost saving is a compact, space-efficient seating plan because this minimizes the total sound absorption of the audience and thus the room volume needed for the specified reverberation time.

The interior shaping of wall and ceiling surfaces (including balcony overhangs) establishes the acoustical character of the space: clarity and loudness of unamplified sounds, balance between sound sources (e.g., sections of orchestra or opera singer relative to pit orchestra), envelopment of sound, and so forth. Detailed analysis of the many variables for a large performance space is complicated, but the basic requirements are straightforward. All potentially useful sound reflecting surfaces should be as hard as possible, located and oriented to distribute a balanced sequence of reflected sounds to each listener. A very short time interval (ideally less than 0.03 s) should occur between the arrival of the "direct" (i.e., unreflected) sound and the first reflections at the listener's ears. Surfaces that would focus sound reflections or return long-delayed echoes (over 0.05 s of delay) to any listener should be reoriented to provide useful reflections. In extreme cases, surfaces that create echoes could be controlled by covering with sound-absorbing material. However, since this added absorption robs the space of reverberation and could necessitate a compensating increase in volume, interior shaping for sound distribution should be the primary goal.

Well-understood techniques allow a performing arts hall to be adapted acoustically to accommodate from a large-scale orchestra plus organ and chorus to musicals and opera within a limited time. Movable sound reflectors, including a demountable orchestra enclosure of variable size, modify the strength and balance of sound reflections; the orchestra pit lift can either serve as a thrust stage, create an orchestra pit, or add prime audience seating; electrically operated sound-absorbing curtains change reverberation characteristics to suit a particular work or to deaden the response of the empty hall during rehearsals. If done effectively, this added flexibility may be the difference between a fixed concert hall with limited use and a multiuse space that will be used many more times per year.

Finally, a high-quality sound amplification system designed to complement the natural acoustics of the hall will provide clarity and intelligibility for a soloist or speaker. It can also accommodate a Broadway musical, including an infrared broadcast system for listeners with impaired hearing, playback of recorded speech or music, and sound effects from any location.

6.5 Worship

Traditional forms of worship have comprised one or more of the following: quiet prayer and meditation, sermon and discussion, choral and instrumental (frequently organ) music. Worship spaces range from small chapels to large cathedrals, many of which have borrowed the complex forms of the Romanesque or Gothic church. The use of sophisticated sound amplification systems for music as well as speech, and the trend to very large congregations, has led to spaces seating 7000–10,000 people in a roughly semicircular "arena" configuration. Extreme examples of this form are too large for serious consideration of natural acoustics and should be considered a special case of "arena" design.

For more conventional sizes and programs, the acoustical requirements for a worship space are similar to those for a performing arts facility (Table 9): well-defined locations of celebrants and congregation, wide range of reverberation requirements, the need for speech clarity, with the added requirement of congregational singing. However, lay readers are seldom trained speakers, and liturgical constraints frequently preclude the use of movable elements or acoustically appropriate wall and ceiling shaping.

Most worship needs can be satisfied fairly well by designing the space to meet music requirements (i.e., long reverberation time) and designing the sound amplification system to ensure intelligibility of speech. If the congregation can afford to install a pipe organ, particular emphasis should be placed on a long reverberation time, enhancement of sound projection from choir and organ, and acoustical support of congregational singing. The location of the organ and shaping of walls and ceiling are particularly important for projection of organ sound and for maximizing the acoustical response of the sanc-

TABLE 9 Acoustical Requirements for Worship Spaces

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Speech	4	2	3	3	1	—	2	Y
Music	4	2	3	3	1	—	5	Y
Choir rehearsal	4	1	2	3	1	—	1/2	—

[a] See Table 1.
[b] See Table 4.

tuary. These conditions entail a large volume, the use of heavy construction materials to minimize absorption of bass sounds, and the avoidance or restricted use of carpet. If music quality is considered less important, adequate loudness and good sound coverage for speech and amplified music can be attained in a light frame construction of modest volume with carpeting and upholstered pews.

Since most worship facilities tend to rely upon natural lighting and ventilation, the choice of building site is of particular importance in ensuring a quiet background for worship. If a completely quiet site is not feasible, the worship space should preferably be shielded from traffic by other buildings. If not, it should be mechanically ventilated, with as few openings as possible facing toward exterior noise sources.

6.6 Industrial/Transport

Acoustical requirements for industrial and transport spaces are presented in Table 10.

Industrial Industrial machinery may create high noise levels but often cannot be enclosed because of the need for accessibility. Sound-absorbing materials on walls and ceiling provide little or no benefit for workers close to the machinery, so special worker enclosures may be needed for hearing protection and voice communication. Possible alternatives include modifications to machinery, rotation of personnel to limit noise exposure, and use of special hearing protectors.

In general, the noisiest items of equipment should be separated and enclosed to limit worker exposure, with extensive use of sound-absorbing materials on ceiling and walls.

Transport Either air or ground transport systems may create serious noise problems for neighboring areas. Initial planning studies should include evaluation of compatibility and may entail preparation of an Environmental Impact Report (EIR) (see Chapter 89). An extreme example is a medical emergency helicopter, which typically

TABLE 10 Acoustical Requirements for Industrial/Transport Spaces

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Factory								
Light industrial	1	1	1	4	3	4	—	Y
Heavy industrial	1	1	1	5	5	4	—	Y
Transport								
Terminal	1	1	1	5	5	3	—	—
Waiting	2	1	1	2	1	3	—	Y
Ticketing	2	1	1	2	1	3	—	Y
Hospital Heli pad	1	1	1	5	5			

[a] See Table 1.
[b] See Table 4.

creates a severe noise impact for both hospital patients and neighboring residential areas. Truck loading areas may be less noisy but could be in use much longer, and at any hour unless time limits are imposed.

Control of noise within terminals, for passengers and employees of transport systems, typically requires acoustically treated walls and ceilings to make waiting areas and workspaces suitable for telephone use and listening to amplified announcements.

6.7 Arena/Stadium

Enclosed arenas and stadiums are too large for unamplified speech or music such as orchestral or choral concerts (Table 11). A basic requirement is sound-absorbing wall and ceiling material for control of reverberation, crowd noise and echoes (long-delayed sound reflections). Design of ventilation systems should be reviewed carefully to ensure that location of fans, layout of diffusers, and supply air velocities are compatible with the desired background noise level.

Amplification of sound requires highly specialized systems that can accommodate a wide range of events and audience sizes. Outdoor stadiums and tent enclosures require added consideration of noise intrusion—if close to an airport, for example—and of the potential for disturbing residential areas up to several miles from the stadium.

TABLE 11 Acoustical Requirements for Arenas and Stadiums

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Arena	2	1	1	5	4		—[c]	Y
Enclosed stadium	1	1	1	5	4		—[c]	Y
Outdoor stadium	1	1	1	5	4		—[c]	Y

[a] See Table 1.
[b] See Table 4.
[c] Special evaluation needed.

TABLE 12 Acoustical Requirements for Critical Uses

	Requirements for Noise Control[a]					Requirements for Interior Design[b]		
						Reverberation Time[c]		
Use of Space	Desired Background	Vibration Sensitivity	Speech Privacy	Maximum Sound	Maximum Vibration	Nonprojection	Projection	Sound Amplification
Dance rehearsal	3	1	2	5	4	—	1/2	Y
Music/dance teaching	2	2	3	5	4	—	1/2	Y
Indoor athletics/aerobics/swimming pool	1	1	1	4	4	4	—	Y
Building in high-noise environment	2	1	2	4	5	4	—	—[c]
Acoustic/vibration-sensitive space	5	4	4	3	1	5	—	—
Acoustic labs/Special standards	5	5	4	4	1	5	—	—
Recording/broadcast	5	3	5	5	1	5	—	—[c]

[a] See Table 1.
[b] See Table 4.
[c] Special evaluation needed.

6.8 Acoustical Requirements for Critical Uses

This category encompasses facilities that must be fully isolated from any intruding noise or vibration as well as facilities with very high noise or vibration levels (Table 12). They typically require highly specialized construction, complete structural separation from other facilities, and possibly restricted access. Particular care is needed to ensure that criteria for acceptable use are clearly defined, and that the best plan locations are selected.

Examples of acoustically sensitive spaces include broadcast/recording studios and anechoic chambers. Vibration-sensitive uses include optical systems with high magnification (e.g., electron microscopes) and manufacturing facilities for computer chips or other high-precision systems.

Facilities with very high noise levels must comply with standards such as established in the United States by OSHA. Examples range from band rehearsal rooms to engine test facilities and shooting ranges. High vibration levels should be anticipated from aerobics and dance rehearsal, indoor running tracks, and heavy machinery such as printing presses, air conditioning equipment, and so forth.

7 ACOUSTICAL ANALYSIS OF SPACE

In this section a worksheet for acoustical analysis of spaces is provided.

Project ______________________________ Title/No. of Space ______________________________

Control of Noise and Vibration

1. Summary of noise control requirements (from space summaries in Section 6)

Desired background	Vibration sensitivity	Speech privacy	Code requirements	Maximum sound level in space	Maximum vibration in space

2. Exterior noise control

Noise sources ______________________________
____
____

Intruding noise (Table 2, a)	Construction alternatives (Table 3)				Judgment factor	Final selection	
	Walls		Roof				
	a	b	a	b		Walls	Roof

3. Interior noise control

Noise sources ______________________________
____
____

Intruding noise (Table 2, b)	Construction alternatives (Table 3)				Judgment factor	Final selection	
	Walls		Floor/ceiling				
	a	b	a	b		Walls	Floor/ceiling

4. Interior vibration control

Vibration sources ______________________________
____
____

Intruding vibration (Table 2, c)	Construction alternatives (Table 3)		Judgment factor	Final selection
	a	b		

5. HVAC system Noise/vibration sources ______________________

____

____

Type of system (Table 2, d)		Construction alternatives (Table 3)		Judgment factor	Final selection
		a	b		
Own					
Other					

6. Other building systems Noise sources ______________________

____

____

Type of System (Table 2, e)	Construction alternatives (Table 3)		Judgment factor	Final selection
	a	b		

Control of Sound within the Space

1. Spaces for nonprojection of sound

Required reverberation time	Acoustical treatment			Sound Amplification/announce system
	Ceiling	Walls	Floor	

2. Spaces for projection of sound

Required reverberation time	Acoustical treatment					Sound amplification system
	Stage	Ceiling	Walls	Floor	Fixed seating	

3. Detailed requirements for auditorium

Seating Capacity		Stage facilities				Adjustable volume	Adjustable absorption
Main floor	Balcony	Pit	Lift	Fly loft	Orchestra shell		

4. Requirements for sound amplification system

Central loudspeaker cluster	Distributed system	Pew-back system	Reverberation enhancement	Sound effects	Audience response

5. Room shaping assumptions for auditorium (see references at end of Chapter 91) (use extra sheets as required)

8 WORK SHEET EXAMPLES

In the following examples of worksheets crit means acoustically critical, special evaluation needed; R means sound reflecting, Y is yes, and N is no:

Control of Noise and Vibration:
Apartment/Condo—Urban

line 1	—	4	1	3	UBC	3	2
line 2	—	4	2/3	2/3	cost	2	2
line 3	—	3	3/4	3/4	cost	3	3
line 4	—	2	3	4	cost	3	
line 5							
own		—	—	—	—	—	
other		1	1	—	—	1	
line 6	—	3	3	—	—	3	

Note—Control of sound within each space not applicable for apartment/condominium.

Control of Noise and Vibration: Multipurpose
Auditorium—Urban

line 1	—	4	3	4	UBC	5	4
line 2	—	5	5/–	5/–	crit	5	5
line 3	—	5	5/4	5/4	crit	5	5
line 4	—	4	4	5	crit	5	
line 5							
own		2	4	5	crit	4	
other		2	4	5	crit	4	
line 6		3	3	5	crit	5	

Control of Sound within the Space

line 1	—	Not applicable						
line 2	—	3	R	R	R	R	Y	Y
line 3	—	1500/750	Y	Y	Y	Y	crit	crit
line 4	—	Y	N	N	N	Y	Y	

97

NOISE CONTROL IN U.S. BUILDING CODES

GREGORY C. TOCCI

1 INTRODUCTION

Noise ordinances, regulations, criteria, and standards are promulgated by organizations to serve the needs of their membership or the public. Cities and towns publish noise ordinances that appear as part of their town regulations or in their zoning bylaws. These usually control environmental sound by focusing on the community, or portions of the community, as a receptor. These are sometimes referred to as *imission* limits, that is, they limit the amount of sound received by a receptor. Many ordinances set explicit imission limits, for example, a maximum allowed sound level at a property line. Imission limits are also sometimes set implicitly, for example, controlling sound in adjacent residential units by setting limits on the sound isolation between residential spaces. Many federal agencies have noise regulations that are intended to serve only a certain group of receptors, for example, only those affected by a specific type of federally funded project.

Emission-type limits set constraints on sound produced by a source. Emission limits are more prevalent among agencies setting standards for specific industries. Often, these are expressed using descriptors other than sound pressure level; for example, sound power level is sometimes used.

This chapter discusses noise limits that appear in model building codes. As discussed, model building code noise limits vary greatly in scope. The following discusses the general structure of building codes and the specific noise limits of each of the three most widely adopted codes. Although these limits must be considered in acoustical design of buildings, they may not adequately serve the needs or meet the expectations of receptors, that is, persons living in a specific building. This chapter concludes with the author's suggestions for establishing criteria that recognize building code limits but that more directly address the expectations of persons living in residential buildings.

Building codes set standards of quality and uniformity for all aspects of building construction. Building codes have been in use since before the 1900s. Since the development of a comprehensive code covering all aspects of building construction exceeded the technical abilities of most cities, even large cities, several organizations were formed between 1920 and 1940 to serve the needs of building regulators in cities and towns. A large part of the work of these organizations was to develop model building codes that could be used by participating localities and states. The three chief building code organizations formed and that represent most U.S. jurisdictions are as follows:

- Building Officials and Code Administrators International (BOCA)
- International Conference of Building Officials (ICBO)
- Southern Building Code Congress International (SBCCI)

Although building codes are intended to be comprehensive, it is only relatively recently that model codes have included limits on noise. This chapter discusses noise limits presented in the model codes developed by these three organizations. Table 3 in Section 5 summarizes the noise limits of these three model codes.

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

2 MODEL BUILDING CODES

All three model code groups are private organizations whose members are local and state building officials responsible for enforcing building codes in their states and localities. The federal government does not require states or localities to have building codes. Federal agencies use their own building codes or use one of the model codes in building their own projects. The three major building code organizations have as their membership building code officials. These organizations serve their membership by producing bodies of code that are adopted by states and localities and administered by the membership. Hence, a large part of the activity of these model code organizations is to educate their membership in code practices and enable them to make judgments regarding the conformance of building construction to code requirements.

Code usage in the United States tends to be regional. States and localities in the northeast and northern part of the Midwest are most often members of BOCA. Those in the western part of the United States most often are members of ICBO. Those in the southern states most often are members of SBCCI. There are many jurisdictions in the United States that do not have building codes. Others have their own building code, for example, New York City and Chicago. Each state determines whether it will adopt the building code on a statewide basis or leave it to the localities to make the decision whether or not to adopt a building code.

3 NOISE REGULATION FORMAT

Generally, noise regulations share several characteristics:

1. *Status.* Every standard states whether it is voluntary or obligatory, and, if it is obligatory, under what circumstances individuals or organizations must conform with the standard.
2. *Object.* Every standard is devised to limit sound exposure at a receiver or to limit sound emitted by a source.
3. *Scope.* Every standard will indicate the specific type of noise exposure and condition under which the ordinance is applicable.
4. *Limit.* Every standard expresses its limits either as numerical values for one or more noise descriptors or in some other specific way such as establishing hours of operation or in some subjective manner such as not allowing noise that constitutes an "annoyance."
5. *Evaluation.* Every standard should indicate the means by which conformance can be objectively established.

4 MODEL BUILDING CODE NOISE LIMITS

The following is a detailed discussion of the noise control provisions of each model code. Further information, or the full code text, can be obtained from each organization.

4.1 BOCA

The Building Officials in Code Administrations International standard for sound appears in Section 714.0, Sound Transmission Control in Residential Buildings. It is part of the body of the BOCA code therefore its status is obligatory; that is, a state or locality that adopts the BOCA building code must include this section on building noise control or pass specific legislation to remove or modify it.

The object of BOCA Section 714.0 is the receiver; that is, it seeks to limit noise at a receptor as opposed to controlling sound emitted by a source. The scope of the code is limited to the sound isolation performance of "all common interior walls, partitions, and floor/ceiling assemblies between adjacent dwelling units or between a dwelling unit and adjacent public areas." Note that this section of the BOCA code only deals with sound transmission between dwelling units even though the BOCA code covers all types of building construction.

The limits of Section 714.0 are sound transmission class (STC) and impact insulation class (IIC) ratings of constructions separating dwelling units. The required wall and floor/ceiling STC rating is 45. The required floor/ceiling IIC rating is 45. A building official responsible for evaluating conformance of residential building partitions and floor/ceiling constructions with BOCA noise provisions can do so by comparing partition and floor constructions with lists of test assemblies provided in the code. Hence, the code does not explicitly require the involvement of an engineer or some other professional and does not require testing. However, the building official may, at his or her discretion, require testing of wall or floor/ceiling constructions not listed in the code or may require testing out of a concern for construction quality.

4.2 ICBO

The International Conference of Building Officials has produced the Uniform Building Code (UBC). Acoustical requirements for buildings appear in UBC Appendix, Chapter 35, Sound Transmission Control. The appendix of the UBC contains only recommendations of ICBO and are not part of the code; thus its status is voluntary. For any provisions of the appendix to become obligatory, a state or locality must adopt all or portions of the appendix into its code. Presently, the State of California has adopted most of the UBC Appendix, including Chapter 35, into its code body. Hence, in California provisions of the Appendix, Chapter 35, Sound Transmission Control, are obligatory.

The object of the UBC Appendix Chapter 35 provisions for sound transmission is the receiver. The scope of Appendix Chapter 35 is to "establish uniform minimum noise insulation performance standards to protect persons within new hotels, motels, dormitories, long-term care facilities, apartment houses and dwellings other than detached single-family dwellings."

The limits of Chapter 35 fall into three categories: sound isolation, impact sound isolation, and allowable interior sound levels. The airborne sound isolation limit is expressed as an STC rating of not less than 50 or a field sound transmission class (FSTC) rating of not less than 45. The FSTC rating is similar to the STC rating except it is applied to noise reduction data collected in the field instead of in a laboratory. Because of limitations associated with building conditions, the FSTC rating of a demising construction is usually 0–5 dB less than the corresponding STC rating determined in a laboratory. This standard, as well as others, recognizes this and permits the FSTC to be up to 5 points less than the corresponding laboratory measured STC rating.

Similarly the IIC rating must not be less than 45. The standard allows the measurement of the field impact insulation class (FIIC) rating but indicates that it must also not be less than 45. The aforementioned discrepancy between laboratory- and field-measured STC ratings does not occur in the impact insulation test. Hence, the laboratory- and field-measured IIC rating values should be in close agreement. The chief difference between the IIC rating and the FIIC rating relates to absorption in the receiving room and the number of measurements and the placement of the tapping machine on the floor tested. Note that STC, FSTC, and IIC ratings are described in American Society for Testing and Materials (ASTM) standards; however, FIIC is not defined in the ASTM standard. Moreover, the Appendix Chapter 35 only indicates that FIIC is the measurement of impact insulation class in the field.

The UBC limit on interior sound levels is expressed as a *day–night average sound level* (DNL) not to exceed 45 dB. The standard permits the alternative use of the *community noise exposure level* (CNEL) and indicates that the interior CNEL sound level must not exceed 45 dB. The day–night average sound level is the 24-h equivalent sound level with a 10-dB penalty added to sound levels occurring between 10 p.m. and 7 a.m. The community noise exposure level is similar in that it is a 24-h equivalent sound level, but with a 5-dB penalty applied to sound levels occurring between 6 p.m. and 10 p.m., and a 10-dB penalty applied to sound levels occurring between 10 p.m. and 7 a.m. The UBC standard encourages the use of day–night average sound level in lieu of the community noise exposure level but indicates that a project should be evaluated by either one or the other, but not both.

A building official responsible for evaluating the conformance of a building with UBC Appendix Chapter 35 sound limits may do so by comparing wall and floor/ceiling constructions with those assemblies listed in UBC Standard No. 35-1, 35-2, or 35-3. Or the official can request certification through field testing by an acoustician. The acoustician can judge conformance or nonconformance of a building construction with the UBC Appendix Chapter 35 standard through the use of one or more of the following noise descriptors:

- Noise isolation class (NIC)
- Normalized noise isolation class (NNIC)
- Field impact insulation class (FIIC)
- Normalized A-weighted sound level difference (D_n)

The NIC and NNIC are defined in ASTM standards. The FIIC and D_n are described in UBC Appendix Chapter 35.

The Appendix Chapter 35 evaluation of interior sound levels requires an acoustical analysis to be conducted before construction if the DNL or the CNEL exceeds 60 dB. The standard also alternatively requires an evaluation report by an acoustician after construction is completed.

4.3 SBCCI

The Southern Building Code Congress International promulgates a separate document as its standard for noise control. It is composed of five chapters. Chapters 1 and 2 are the introduction and a description of terms. Chapter 3, "Noise Disturbances," sets limits on source noise emissions. Chapter 4, "Airport Noise," sets limits on aircraft noise at the receiver (imissions). Chapter 5, "Sound Isolation in Multi-family Dwellings," sets limits on partition sound isolation performance. The entire standard is voluntary and is only a recommendation of SBCCI for those communities wishing to implement a noise regulation.

SBCCI Chapter 3: Noise Disturbance The scope of SBCCI Chapter 3 indicates that it "shall apply to any sound producing source referenced in this standard or any sound producing source that can effect the health, safety, and comfort of those persons within audible range of the sound."

The limits set in SBCCI Chapter 3 are varied. They include sound pressure level limits for certain types of sources, for example, air compressors, motor vehicles, and so on. For animals and motor boats, subjective limits are set. For construction, materials handling, and so on, time limits are set. This chapter also defines geographical limits, that is, noise-sensitive zones or zoning districts.

The evaluation process for SBCCI Chapter 3 includes permits for certain types of activities. It also establishes a Board of Sound Control, composed of five members including one architect or professional engineer, one attorney, one general contractor, two members-at-large, and a nonvoting "sound control official" who is responsible for evaluating acoustical conditions and implementing decisions of the Board of Sound Control. This requirement to establish a board of sound control is unique among model codes.

SBCCI Chapter 4: Airport Noise The scope of the SBCCI Chapter 4 standard for airport noise defines "an Airport Noise Impact Overlay District for the purpose of controlling conflicts between land uses and noise generated by aircraft, ... and ... establishes sound isolation requirements for exterior walls and roofs of buildings located in the Airport Noise Impact Overlay District."

The limits used in this standard are given in Table 1. The table indicates minimum required day–night average sound level noise reductions of exterior building facades. Minimum required noise reductions depend on exterior sound level (expressed as ranges of DNL) and depend on building use. This table is very much like the table used by the Federal Aviation Administration (FAA) in 14 CFR Part 150 for airport land-use compatibility studies.

The evaluation procedure involves certification by a registered professional architect or engineer that the exterior walls, roofs, and windows conform with the minimum STC ratings specified in Table 2.

To use this table for a specific building the following procedure should be followed:

- First, determine the appropriate range of opening percentage (percentage of the total wall/ceiling occupied by fenestration) in the first column.
- Then, locate the corresponding required minimum noise reduction (determined from Table 2) in the second column.
- For this method to be valid, the exterior walls and roofs must have the minimum STC ratings specified in the third column.
- The required fenestration (exterior doors, windows, and sloped glazing) should have the minimum STC rating specified in the fourth column.

If, in the unlikely event, the exterior wall of a building has an STC rating less than specified in the third column

TABLE 1 Noise Level Standards

	Exterior Day–Night Average Sound Level (dB) by Occupancy			
Building Type	75+	70–75	65–70	65–
Group A	NP	30	25	NSR
Public assembly halls, restaurants, motion picture theaters, churches, museums, heaters for stage production, auditoria				
Other	NP	NSR	NSR	NSR
Group B				
Libraries	NP	30	25	NSR
Other	30	25	NSR	NSR
Groups E and I	NP	30	25	NSR
Group M	30	25	NSR	NSR
Group R				
Hotels and motels	35	30	25	NSR
Other	NP	30	25	NSR
Other	NSR	NSR	NSR	NSR

Notes: NP, not permitted; NSR, no special requirement.

Source: Table 403.1 of SBCCI Chapter 4.

TABLE 2 Minimum Sound Transmission of Assemblies

Percent Openings	Minimum Required Noise Reduction (DNL) in dB	Minimum TL of Exterior Walls and Roofs	Minimum of Exterior Doors, Windows, and Sloped Glazing
1–25	35	50	42
	30	45	37
	25	39	28
26–70	35	55	45
	30	50	41
	25	45	37

Source: Table 403.2 of SBCCI Chapter 4.

of Table 2, then a higher STC rating than shown in the fourth column may be needed. It should be noted, however, that the STC ratings given in the third column are lower than the STC ratings for most standard exterior wall constructions, so that the fenestration STC ratings required in the fourth column are conservatively high. In lieu of using Table 2, some economy in the selection of windows can be obtained by an acoustical analysis during building design. This can often result in fenestration STC ratings that are less than those shown in the fourth column of Table 2 while still achieving the required noise reduction in Table 1.

The purpose of Table 2 is to assure that acceptable STC ratings are used in the event that a more sophisticated analysis by an acoustical consultant is not available. The risk to a building owner using this table is that it may require a window sound isolation performance that is greater than actually needed to achieve the noise reductions of Table 1.

SBCCI Chapter 5: Sound Isolation in Multi-Family Dwellings For its scope, SBCCI Chapter 5 indicates that "this chapter shall apply to the minimum allowable sound isolation for partitions that separate adjacent units in multi-family dwellings, and similar partitions that separate a dwelling unit from public areas, service areas, or commercial facilities."

Chapter 5 of SBCCI limits partition sound isolation between adjacent dwelling units or between dwellings units and public areas such as hallways, service areas, or commercial facilities, to a minimum STC rating of 45.

Unlike other chapters of the SBCCI standard for sound control, the evaluation process to be undertaken by a building official is not clearly established. It does imply, however, that several publications may be used for reference. These present STC ratings for a variety of partition constructions. Those having a minimum STC rating of 45 can be judged acceptable.

5 COMPARISON OF MODEL BUILDING CODES

Table 3 summarizes the provisions of the three model codes by briefly indicating status, object, scope, limits, and evaluation procedures for each model code.

6 BUILDING CODES OUTSIDE THE UNITED STATES

Many countries around the world maintain uniform building codes with provisions controlling sound in buildings. These use a variety of descriptors, many specifically devised for the code in which they are cited, and others widely used in national and international (ISO) standards. Tables A1–A10 in the Appendix present code requirements for 10 countries: Austria, Denmark, Estonia, France, Korea, The Netherlands, Germany, Russia, Sweden, and the United Kingdom.

7 GUIDE TO BUILDING NOISE CRITERIA

Conformance with building code requirements for sound isolation does not always ensure that the expectations of occupants of a building will be met. When providing assistance to an architect or owner of a new building, the acoustical consultant must recognize this and be vigilant in identifying those situations where occupants will expect greater building sound isolation performance. It is this author's experience that often buildings in which sound isolation complaints have occurred, have had sound isolation characteristics conforming with minimum code requirements. Perhaps the most common instance of complaints is one where a family sells a single-family home in a quiet suburb and moves to a more expensive condominium, many times in a more urban setting. Having been swayed by their perceived

TABLE 3 Summary of Model Noise Codes

BOCA	ICBO	SBCCI
	Title	
Section 714.0: Sound Transmission Control in Residential Buildings	Appendix Chapter 35: Sound Transmission Control	Standard for Sound Control Ch. 3: Noise Disturbances Ch. 4: Airport Noise Ch. 5: Sound Isolation in Multi-Family Dwellings
	Status	
Obligatory	Voluntary (obligatory in California)	Voluntary
	Object	
Receiver	Receiver	Ch. 3: Source Chs. 4, 5: Receiver
	Scope	
Interior walls, and floors between dwelling units	Interior walls, floors between dwelling units, exterior walls Interior sound levels produced by exterior noise sources	Ch. 3: Exterior noise sources Ch. 4: Exterior wall, windows, doors of buildings near airports Ch. 5: Interior walls, floors between dwelling units
	Limits	
STC ≥ 45 IIC ≥ 45	STC ≥ 50 (FSTC ≥ 45) IIC ≥ 45 (FIIC ≥ 45) DNL or CNEL ≤ 45 dB (Appendix to CA Amendment)	Ch. 3: Various Ch. 4: Exterior wall/window STC depending on exterior DNL Ch. 5: STC ≥ 45
	Evaluation	
Comparison with listed assemblies	Comparison with listed assemblies and report by acoustical consultant (testing if needed)	Ch. 3: Board of Sound Control Ch. 4: Certification by P. E. or Registered Architect, conformance with specified values Ch. 5: Comparison with listed assemblies

quality of the condominium's fit and finish, as well as the unit's cost, and having lived for many years in a single-family home, the reality of hearing sounds from adjacent units and/or mechanical equipment sound often becomes a source of disappointment and complaints. The acoustical consultant must guide the architect in avoiding this pitfall by providing recommended sound isolation performances that meet the likely expectations of future occupants while fitting within the reality of budget constraints.

Table 4 presents a guide for establishing sound isolation criteria for three categories of residential construction defined as minimally sensitive, moderately sensitive, and highly sensitive. These sensitivities are of persons occupying multiresidential buildings. Persons falling in the minimally sensitive category in this context are those not expecting special qualities in building performance or do not anticipate intensive use of their living unit for activities requiring quiet, for example, studying or intently listening to quiet music. High ambient sound levels produced by mechanical systems are typical of living units served by window air conditioners. High ambient sound levels are typical of living units in buildings overlooking major highways and with windows providing low noise reduction because of lightweight glazing and/or because of poor window gaskets.

Persons falling in the moderately sensitive category would expect good speech privacy and low enough ambi-

TABLE 4 Guidelines for Acoustical Criteria in Multiresidential Buildings

Sound Isolation Type	Minimally Sensitive	Moderately Sensitive	Highly Sensitive
Party wall/floor sound isolation			
Wall and floor sound isolation (STC rating)	45	55	65
Floor impact sound isolation (IIC rating)	45	55	65
Mechanical equipment sound isolation			
Sleeping rooms (Criteria cruve, NC, RC)	40	30	20
All other rooms (Criteria curve, NC, RC)	45	35	25
Exterior environmental sound isolation			
Exterior environmental sound levels inside units (day–night average sound level in dB)	45	35	25
Maximum exterior environmental sound level inside unit [A-weight sound pressure level in dB(A), slow response]	65	55	45

ent sound levels to ensure uninterrupted intelligibility of spoken and television speech at moderate levels. Mechanical systems serving these units would require sound attenuation. In noisier urban areas, acoustical windows may be required to sufficiently control exterior sound transmission into living areas. Achieving the IIC ratings in Table 4 would require standard carpeting or faily sophisticated hard-finished floor systems incorporating resilient underlayments.

Persons falling in the highly sensitive category would expect minimally audible airborne and impact sound transmission from adjacent units, good speech intelligibility for low-level signals, and an ability of playing moderately loud music without disturbing neighbors. Even in quiet environs, acoustical windows may be required. Special attention must also be paid to controlling noise from mechanical systems serving these living units. Achieving the IIC rating indicated in Table 4 would require carpeting or floated hard-finished floors. Floated floors would be characterized as a concrete or other heavy construction isolated from the building structural floor by vibration isolation elements.

To achieve building floor/ceiling constructions that provide IIC ratings of 50 or higher generally requires special construction details including resilient floor underlayments (or floated floor constructions) and sound isolation ceilings. IIC ratings higher than 65 are generally easily obtained by using floor carpeting. As a compromise, condominium complexes sometimes create rules that permit hard finished floors, providing they can achieve a minimum IIC rating of 50, if carpeting is used in unit walkways and halls, and if the total carpeting floor coverage is in a range of 40–70%.

Interior sound levels produced by mechanical systems and indicated in Table 4 are consistent with the recommendations of the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE).[1] Tonal sound, such as that produced by transformers and other electrical equipment, may need to adhere to lower sound level limits than indicated for mechanical systems. This is also true for transient sound, such as sound produced by elevators.

The suggested criteria for environmental sound in Table 4 apply to transportation sound, that is, sound produced by road traffic, aircraft, and railroads. The criteria may not be suitable for industrial noise or weapons firing sound, for example.

The criteria of Table 4 are to be used only as a general guide for setting criteria and should not be included in building project specifications without further consideration of project needs and code requirements. The interpretation of sensitivity for categorizing residential occupanices must be done with care, and individual circumstances may influence categorizations.

8 SUMMARY

The BOCA Section 714.0, Sound Transmission Control in Residential Buildings, sets limits on the sound isolation between dwellings. This code does not address exterior sound or sound transmitted into buildings from the exterior.

The ICBO Appendix Chapter 35 of the UBC sets limits on both interior sound levels and exterior building facade sound isolation performance. The exterior facade sound isolation limits are equivalent to those recommended in FAA 24 CFR, Part 150. The standard provides a comprehensive evaluation process for the building official and makes use of the expertise of an acoustician.

The SBCCI Standard for Sound Control sets limits on sound produced by various sources, sets limits on exterior building facade sound isolation performance, and sets party wall sound isolation limits. The standard is voluntary and involves a comprehensive evaluation process.

From the standpoint of acoustical engineering, the ICBO noise standard in the UBC appears to be the best organized and most practical to implement but covers only residential buildings. The SBCCI noise standard is somewhat more comprehensive in that it covers more building types. Portions of the SBCCI standard requiring a board of sound control and a sound control officer would be difficult to implement. The BOCA noise standard, although restricted to a single issue, which is sound isolation between dwellings, is part of the code; that is, it is obligatory.

In general, all of the sound transmission class rating requirements between dwellings specified in the model codes are, in the author's experience, insufficient to avoid common complaints. It is not unusual for acoustical consultants to recommend higher limits for the design of multifamily dwellings. Table 4 provides suggested criteria for multiresidential buildings according occupant expectations.

In spite of the shortcomings cited, noise control requirements limiting the transmission of sound into buildings and between dwelling units are crucial. Programs to establish compatible land-use zoning around airports and other noise producing activities have been widely established. Nevertheless, economic and social pressures are still causing noise-sensitive building development to occur near major noise sources. Today, construction of residential subdivisions within a few feet of major highways and in areas affected by aircraft noise is still common. Oddly enough, sometimes the attraction of noisy sites for development is the source itself, that is, the convenience of being near a highway or an airport. Hence, the only way to protect the eventual building occupant is for jurisdictions to incorporate sound isolation provisions into their building codes.

Building codes outside of the United States address the same issues but with a wide variety of descriptors. Most uniformity among the codes presented in the Appendix to this chapter is among requirements for sound isolation between dwellings. These also tend to be more demanding than U.S. building codes.

Acknowledgments

The author acknowledges Birgit Rasmussen, The Acoustics Laboratory of the Technical University of Denmark, for her work in collecting the acoustical standards of the 10 countries presented in the Appendix to this chapter. The author also thanks Brion G. Koning, Cavanaugh Tocci Associates, for his assistance in gathering information for this and Chapter 16 of Volume 2. Finally, the author thanks Randolph E. Wright and Monsanto for their support in gathering U.S. building code information together.

REFERENCE

1. "Sound and Vibration Control," *1995 ASHRAE Applications Handbook*, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Chapter 43, p. 43.5.

APPENDIX: TABLES OF ACOUSTIC CODES

Tables A1–A10 in this Appendix provide acoustical conditions for 10 countries as of March 1994.* An overview of the main requirements is presented for each country.

---

*The data in this Appendix are from B. Rasmussen and J. H. Rindel, "Lydforhold I Boliger: State-of-the-Art," The Acoustics Laboratory, Technical University of Denmark, Lyngby, Denmark, September 1994.

TABLE A1 Austria

Requirement[a]	Symbol and Value [1][a]		Comments[b]
	Flats	Row Houses	
Airborne sound insulation	$D_{nT,w}$	$D_{nT,w}$	Test and rating method: ÖNORM S 5100 Teil 1[c]
Between dwellings (H = horizontal V = Vertical)	H: ≥55 dB V: ≥55 dB	≥60 dB	
Impact sound insulation	$L'_{nT,w}$	$L'_{nT,w}$	Test and rating method: ÖNORM S 5100 Teil 2[c]
Between dwellings	≤48 dB	≤46 dB	
Installation noise	L_A	L_A	Test method: ÖNORM S 5102[c]
Joint water installations, heating installations, and technical services	≤25 dB	≤25 dB	For continuous or intermittent noise (e.g., pumps)
	≤30 dB	≤30 dB	For short noises (e.g., lifts, toilet-flashing)
Traffic noise	$R'_{\text{res},w}$	$R'_{\text{res},w}$	Sound insulation of the facade, including, e.g., windows
Roads, railways, aircaft; $L_{A,\text{eq}}$ in front of facade: Day–night (dB)			Test and rating method: ÖNORM S 5100 Teil 3[d]
≤55/ ≤45	33 dB	33 dB	
56–65/46–55	38 dB	38 dB	
66–70/56–60	43 dB	43 dB	
71–75/61–65	48 dB	48 dB	
76–80/66–70	53 dB	53 dB	
81–85/71–75	58 dB	58 dB	
>85/ >75	63 dB	63 dB	

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publication: [1] ÖNORM B 8115-2 Hochbau, Schallschutz, und Hörsamkeit, Anforderungen an den Schallschutz, November 1992.

[b] For example, concentration, test methods, classification, and trends.

[c] Proposals for increased requirements are found in the publication given in footnote a.

[d] Additional requirements for $L_{A,\max} > 106$ dB (near airports) are found in the publication given in footnote a.

TABLE A2 Denmark

Requirement[a]	Symbol and Value		Comments[b]
	Flats [1][a]	Row Houses [2][a]	
Airborne sound insulation	R'_w	R'_w	Test method: ISO 140/4 Rating method: ISO 717/1
Between dwellings (H = horizontal, V = Vertical)	H: ≥52 dB V: ≥53 dB	≥55 dB	
Impact sound insulation	$L'_{n,w}$	$L'_{n,w}$	Test method: ISO 140/7 Rating method: ISO 717/2
Between dwellings	≤58 dB	≤53 dB	
Installation noise	$L_{A,\text{eq},2\text{min}}$	$L_{A,\text{eq},2\text{min}}$	Unfurnished rooms
Joint water installations	≤35 dB	≤35 dB	No requirement for noise arising from use of own installation
Joint heating and air-conditioning installations	≤30 dB	≤30 dB	
Joint mechanical installations	≤35 dB	≤35 dB	
Traffic noise	$L_{A,\text{eq},24\text{h}}$	$L_{A,\text{eq},24\text{h}}$	Indoor level
Roads	≤30 dB	≤30 dB	
Railways	≤30 dB	≤30 dB	
Aircraft			

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publications: [1] "Bygningsreglement 1982" ("Building Regulations 1982"), Byggestyrelsen 1982; [2] "Bygningsreglement for småhuse 1985" ("Danish Building Regulations for Small Buildings 1985"), Byggestyrelsen 1985. Other requirements: "Ekstern støj fra virksomheder" ("External Noise from Industries"), Vejledning nr. 5/1984, Miljøstyrelsen 1984.

[b] See footnote *b*, Table A1.

TABLE A3 Estonia

	Symbol and Value		
Requirement[a]	Flats	Row Houses	Comments[b]
Airborne sound insulation [1][a]	R'_w	R'_w	Test method: ISO 140/4 Rating method: ISO 717/1 See footnote *c*
Between dwellings (H = horizontal, V = Vertical)	H: ≥52 dB V: ≥52 dB	≥52 dB	
	(V: ≥53 dB)	(≥55 dB)	See footnote[d]
Impact sound insulation [1][a]	$L'_{n,w}$	$L'_{n,w}$	Test method: ISO 140/7 Rating method: ISO 717/2 See footnote *c*
Between dwellings	≤60 dB	≤60 dB	
	(≤58 dB)	(≤58 dB)	See footnote *d*
Installation noise [2][a]	L_A or $L_{A,eq}$	L_A or $L_{A,eq}$	Furnished rooms[e]
Day 07:00–22:00	≤40 dB	≤40 dB	
Night 22:00–07:00	≤30 dB	≤30 dB	
Traffic noise [2][a]	$L_{A,eq}$	$L_{A,eq}$	See footnote *f*
Roads, railways, aircraft			
Day, 07:00–22:00	≤40 dB	≤40 dB	
Night, 22:00–07:00	≤30 dB	≤30 dB	

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publications: [1] "Temporary Regulations of Building Acoustics and Protection Against Noise," Tallinn, 1992; [2] Sanitary standard 3077-84, "Permissible Noise Levels in Dwellings, Public Buildings and Living areas," Moscow, 1984.

[b] See footnote *b*, Table 1.

[c] In addition to the rating method ISO 717, the Estonian regulations prescribe a maximum 8 dB unfavourable deviation.

[d] Values of R'_w and $L'_{n,w}$ in parentheses are recommendations only.

[e] The type of installation noise is not specified. In case of impulsive noise or pure tones, 5 dB should be added to the measured level. The maximum noise level should not exceed the permissible level by more than 15 dB.

[f] The maximum noise level should not exceed the permissible level by more than 15 dB.

TABLE A4 France

	Symbol and Value [1][a]		
Requirement[a]	Flats	Row Houses	Comments[b]
Airborne sound insulation	D_{nAT}	D_{nAT}	Test and rating methods: NF S 31-057 (survey method) NF S 31-054 See footnote *c*
Between dwellings (H = horizontal, V = vertical)	H: ≥51 dB(A) V: ≥51 dB(A)	≥51 dB(A)	See footnote *d*
Impact sound insulation	L_{nAT}	L_{nAT}	Test and rating methods: NS S 31-057 (survey method) NF S 31-056 See footnote *c*
Between dwellings	≤70 dB(A)	≤70 dB(A)	See footnote *d*
Installation noise	L_p	L_p	Test method: NF S 31-057 (Survey method)[c]
Joint water installations	≤30 dB(A)	≤30 dB(A)	
Traffic noise	D_{nAT}	D_{nAT}	Test and rating methods: NF S 31-057 (survey method) NF S 31-055 See footnote *c*
Roads, railways	≥30–45 dB(A)	≥30–45 dB(A)	The exact requirement depends on the category of road or railway track and the exposition of the building
Aircraft	≥35 dB(A)	≥35 dB(A)	Zone C

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publication: [1] "Isolation acoustique dans les batiments d'habitation," *Journal Officiel des* 24.06.69 et 07.01.77.

[b] See footnote *b*, Table A1.

[c] A method for quality labeling of dwellings is found in "Guide Qualitel 1993" (two out of seven criteria are related to the acoustical conditions).

[d] New acoustical regulations will be put into force from January 1996. The new requirements will specify $D_{nAT} \geq 54$ dB(A) and $L_{nAT} \leq 65$ dB(A). In 1999 $L_{nAT} \leq 61$ dB(A) will be required.

TABLE A5 Korea

	Symbol and Value		
Requirement[a]	Flats	Row Houses	Comments[b]
Airborne sound insulation [1, 2][a]	D	D	Test method: KS F 2809[c,d]
Between dwellings			
125 Hz	≥30 dB	≥30 dB	
500 Hz	≥45 dB	≥45 dB	
2 kHz	≥55 dB	≥55 dB	
Impact sound insulation [2][a]	L	L	Test method: KS F 2810
Between dwellings			See footnotes *e*, *f*, *g*
Lightweight source	≤L-70	≤L-70	
Heavy-weight source	≤L-50	≤L-50	
Installation noise [2][a]	$L_{A,\max}$	$L_{A,\max}$	Test method: KS A 0701
Joint water installations	≤40 dB	≤40 dB	See footnotes *d*, *e*, *f*
Traffic noise [2][a]	L_{dn}	L_{dn}	Test method: notification of the Ministry of Construction 463 (1986)
Roads, railways, aircraft	≤40 dB	≤40 dB	See footnotes *d*, *f*

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publications: [1] Building Act and Regulation, 1990; [2] Housing Control Act and Regulation, 1987.

[b] See footnote *b*, Table A1.

[c] Notification of the Ministry of Construction 341 (1990.6).

[d] Changed requirements are expected before the end of 1994.

[e] These requirements are for industrialized housing.

[f] These requirements are for Korean National Housing Corporation.

[e] The test and rating method for impact sound insulation are very similar to the Japanese standards JIS A 1418 and JIS A 1419, but requirements are different.

TABLE A6 The Netherlands

	Symbol and Value [1][a]		
Requirement	Flats	Row Houses	Comments[b]
Airborne sound insulation	$I_{lu;k}$	$I_{lu,k}$	Test and rating method: NEN 5077[c,d e]
Between dwellings (H = horizontal, V = Vertical)	H: ≥0 dB V: ≥0 dB	≥0 dB	
Impact sound insulation	I_{co}	I_{co}	Test and rating method: NEN 5077[d,e,f]
Between dwellings	≥0 dB	≥0 dB	Floor without covering
Installation noise	$L_{IA;K}$	$L_{IA;k}$	Test and rating method: NEN 5077[d,e,g]
Joint water installations	≤30 dB	≤30 dB	No requirement for noise arising from use of own installations
Heating installation	≤30 dB	≤30 dB	
Ventilation systems	≤30 dB	≤30 dB	
Elevators	≤30 dB	≤30 dB	
Traffic noise	$G_{A;k}$	$G_{A;k}$	Test and rating method: NEN 5077[d,e,h]
Roads, railways			Criterium: Indoor level ≤35, 30, 25 dB(A) for day, evening, night, respectively
Aircraft			Roughly as above

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publications: [1] "Bouwbesluit" ("Building Regulations"), December 1991.

[b] See footnote *b*, Table A1.

[c] $I_{lu;k} \approx R'_w - 55$ dB.

[d] All requirements relate to the combination of rooms on a floor level. For further details, see NEN 5077.

[e] A definition of quality classes for rooms and dwellings is under preparation (expected in 1995). The classification will be based on a single-number rating complying with the revised ISO 717 (expected 1995).

[f] $I_{co} \approx 70 - L'_{nT,w}$ decibels for bare concrete floors or $I_{co} \approx 59 - L'_{nT,w}$ decibels for other floors, such as wooden floors, floating floors, and floors with soft coverings.

[g] Measurement based on $L_{A,\max,\text{slow}}$.

[h] $G_{A;k} \approx R'_{A,\text{tr}} - 3$ dB. The required $G_{A;k}$ depends on the outdoor sound level.

TABLE A7 Germany

Requirement[a]	Symbol and Value [1][a]		Comments[b]
	Flats	Row Houses	
Airborne sound insulation	R'_w	R'_w	Test method: DIN 52210-3 Rating method: DIN 52210-4 See footnotes *c*, *d*
Between dwellings (H = horizontal, V = Vertical)	H: ≥53 dB V: ≥54 dB	≥57 dB	
Impact sound insulation	$L'_{n,w}$	$L'_{n,w}$	Test method: DIN 52210-3 Rating method: DIN 52210-4 See footnotes *c*, *d*
Between dwellings	≤53 dB	≤48 dB	
From stairways, etc.	≤58 dB	≤53 dB	
Installation noise			Test method: DIN 52219 See footnotes *c*, *d*
Joint water installations	$L_{In} \le 35$ dB	$L_{In} \le 35$ dB	No requirement for noise arising from use of own installation
Other technical services	$L_{AF,\max} \le 30$ dB	$L_{AF,\max} \le 30$ dB	
Traffic noise	$R'_{res,,w}$	$R'_{res,w}$	Sound insulation of the facade, incl. windows etc.
Roads, railways, aircraft; $L_{A,eq}$ in front of the facade: day (dB)			Test method: DIN 52210-5 Rating method: DIN 52210-4 See footnote *d*
≤55	30 dB	30 dB	The same requirements are valid for outdoor industrial noise
56–60	30 dB	30 dB	
61–65	35 dB	35 dB	
66–70	40 dB	40 dB	
71–75	45 dB	45 dB	
76–80	50 dB	50 dB	
>80	—[e]	—[e]	

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publication: [1] "DIN 4109—Schallschutz im Hochbau. Anforderungen und Nachweise," November 1989.

[b] See footnote *b*, Table A1.

[c] Recommendations for increased requirements are found in Beiblatt 2 zu DIN 4109.

[d] A proposal for classification (three classes) of dwellings is found in VDI 4100 Entwurf, October 1989 (Noise Control in Housing—Criteria for Planning and Assessment). The final version of VDI 4100 is expected in the middle of 1994. Class I will correspond to DIN 4109. Stricter values will be defined for the two other classes, corresponding to a higher comfort.

[e] Depends on local conditions.

TABLE A8 Russia

Requirement[a]	Symbol and Value		Comments[b]
	Flats	Row Houses	
Airborne sound insulation [1][a]	I_b		Test method: GOST 27296-87 Rating method: SNIP II-12-77 Part 12
Between dwellings (H = horizontal, V = Vertical)	H: ≥50 dB V: ≥50 dB		See footnote *c* See footnote *d*
Impact sound insulation [1][a]	I_y		Test method: GOST 27296-87 Rating method:
Between dwellings	≤67 dB		SNIP II-12-77 Part 12 See footnote *e* See footnote *d*
Installation noise [2][a]	L_A or $L_{A,\text{eq}}$		Furnished rooms See footnotes *d*, *f*
Day, 07:00–22:00	≤40 dB		
Night, 22:00–07:00	≤30 dB		
Traffic noise [2][a]	$L_{A,\text{eq}}$		See footnotes *d*, *g*
Roads, railways, aircraft			
Day, 07:00–22:00	≤40 dB		
Night, 22:00–07:00	≤30 dB		

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publications: [1] "SNIP II-12-77 Part 12, Protection against noise," GOSSTROI, Moscow 1987; [2] Sanitary standard 3077-84, "Permissible Noise Levels in Dwellings, Public Buildings and Living Areas," Moscow 1984.

[b] See footnote *b*, Table A1.

[c] $I_b = R'_w - 2$ dB.

[d] No requirements for now houses. Probably the requirements for flats are used.

[e] $I_y = L'_{n,w} + 7$ dB.

[f] The type of installation noise is not specified. In case of impulsive noise or pure tones, 5 dB should be added to the measured level. The maximum noise level should not exceed the permissible level by more than 15 dB.

[g] The maximum noise level should not exceed the permissible level by more than 15 dB.

TABLE A9 Sweden

Requirement[a]	Symbol and Value [1][a]		Comments[b]
	Flats	Row Houses	
Airborne sound insulation	R'_w	R'_w	Test method: SS 02 52 54 Rating method: SS-ISO 717/1 See footnotes *c*, *d*
Between dwellings (H = horizontal, V = vertical)	H: ≥52 dB V: ≥53 dB	≥55 dB	
Impact sound insulation	$L'_{n,w}$	$L'_{n,w}$	Test method: SS 02 52 54 Rating method: SS-ISO 717/2 See footnotes *c*, *d*
Between dwellings	≤58 dB	≤58 dB	
From stiarcases etc.	≤64 dB	≤64 dB	
Installation noise			Test method: SS 02 52 53 Furnished rooms See footnote *d*
Joint water installations	$L_{A,\max} \le 35$ dB	$L_{A,\max} \le 35$ dB	No requirement for noise arising from use of own installation
Joint heating and air conditioning installations	$L_{A,\text{eq}} \le 30$ dB	$L_{A,\text{eq}} \le 30$ dB	In rooms intended for rest and sleep $L_{C,\text{eq}} \le 50$ dB is also required
Traffic noise	$L_{A,\text{eq},24\text{h}}$	$L_{A,\text{eq},24\text{h}}$	Indoor level[e]
Roads	≤30 dB	≤30 dB	See footnote *d*
Railways			
Aircraft			

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publication: [1] "Boverkets Byggregler 94, BBR-94." See footnote *c* below.

[b] See footnote *b*, Table A1.

[c] In addition to the rating method, the Swedish building regulations prescribe a maximum of 8 dB unfavorable deviation.

[d] A proposal for classification of dwellings is found in "Handlingsplan mot buller, SOU 1993:65. Miljö- och naturresursdepartementet.

[e] Recommendation only.

TABLE A10 United Kingdom

Requirement[a]	Symbol and Value [1][a] Flats	Symbol and Value [1][a] Row Houses	Comments[b]
Airborne sound insulation	$D_{nT,w}$	$D_{nT,w}$	Test method: BS 2750 Rating method: BS 5821
Between dwellings (H = horizontal, V = vertical)	H: ≥52 dB V: ≥51 dB	≥52 dB	See footnotes *c*, *d* See footnotes *c*, *e*
Impact sound insulation	$L'_{nT,w}$		Test method: BS 2750 Rating method: BS 5821
Between dwellings	≤62 dB		See footnotes *c*, *f*
Installation noise			
Traffic noise Roads Railways Aircraft			Conditions are set by the local authority based on official guidance [2][a]

[a] The given requirements are only main requirements. These requirements, additional requirements, and further details are found in the following publications: [1] "The Building Regulations. Approved Document E: Resistance to the Passage of Sound," 1992 edition, Department of the Environment and the Welsh Office; [2] "PPG Planning and Noise," to be published by the Department of the Environment during 1994.

[b] See footnote *b*, Table A1.

[c] Acceptable mean values in case of eight or more pairs of rooms tested. Where less pairs (down to one) have been tested, the requirements are 1 dB stricter.

[d] Individual results should not be less than 49 dB.

[e] Individual results should not be less than 48 dB.

[f] Individual results should not be more than 65 dB.

98

NOISE CONTROL FOR MECHANICAL AND VENTILATION SYSTEMS

R. M. HOOVER AND R. H. KEITH

1 INTRODUCTION

This chapter deals with the level and character of sound for the principal types of electrical and mechanical equipment found in the mechanical equipment rooms of typical buildings. Sound pressure (L_p) and sound power (L_w) data for mechanical equipment are presented. Where possible, the data have been correlated with some of the more obvious equipment operational parameters, such as type, speed, power rating, and flow conditions. The levels quoted are suggested for design uses; these levels represent approximately the 80–90 percentile values and assume no extraordinary noise control. Specific information on equipment description and operation is not presented, however, this information is readily available.[1,2]

2 SOUND LEVEL FOR TYPICAL EQUIPMENT

These data are presented in the form of octave band sound power levels (at 10–12 W) or normalized octave band pressure levels. The normalized octave band sound pressure data are representative of a distance at 0.9 m and a room constant of 74 m^2. Sound pressure level data measured at other distances and room constants may be converted to these normalized conditions by using the methodologies found in Chapters 77 and 90. A-weighted sound levels are also given. A-weighted sound levels are useful for comparing the noise output of competitive equipment. For a complete analysis octave band levels should be used.

2.1 Chillers

Chillers are used to provide cooling water for air conditioning systems. The three principal types are (1) packaged chillers with reciprocating compressors, (2) packaged chillers with rotary-screw compressors, and (3) packaged chillers with centrifugal compressors. In addition absorption machines, which are heat-operated refrigeration machines, are sometimes used to provide cooling water. Normalized sound pressure level data for each of these machines are given in Table 1 (Ref. 3, pp. 7-3–7-6). Specific considerations for each classification are as follows:

(a) *Packaged Chillers with Reciprocating Compressors.* The data for these chillers have been reduced to the normalized distance from the acoustical center of the assembly. When cooling requirements exceed 700 kW, centrifugal compressors become more economical, so there are few reciprocating units rated above 700 kW. Although major interest is concentrated here on the compressor component of a refrigeration machine, an electric motor is usually the drive unit for the compressor. The noise levels attributed here to the compressor will encompass the drive motor most of the time.

(b) *Packaged Chillers with Rotary-Screw Compressors.* The octave band sound pressure levels represent near-maximum noise levels for the size range of 350–1000 kW cooling capacity, operating at or near 3600 rpm.

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

TABLE 1 Normalized Sound Pressure Levels, at 0.9 m, for Chillers

Octave Frequency Band (Hz)	Reciprocating Compressors		Rotary-Screw Compressors, 350–1000 kW Cooling Capacity	Centrifugal Compressors		Absorption Machines, All Capacities
	35–175 kW Cooling Capacity	175–700 kW Cooling Capacity		Under 1750 kW Cooling Capacity	1750 kW More Cooling Capacity	
31	79	81	70	92	92	80
63	83	86	76	93	93	82
125	84	87	80	94	94	82
250	85	90	92	95	95	82
500	86	91	89	91	93	82
1000	84	90	85	91	98	81
2000	82	87	80	91	98	78
4000	78	83	75	87	93	75
8000	72	78	73	80	87	70
A-weighted	89	94	90	97	103	86

(c) *Packaged Chillers with Centrifugal Compressors.* Noise levels for packaged chillers with centrifugal compressors are representative for both hermetic and nonhermetic units. These compressors range in size from 351 to 14,067 kW. The noise levels may be influenced by the motors, gears, or turbine drives, but the data provided emphasize the compressor noise. The low-frequency noise levels reflect the increase found for off-peak loads for most centrifugal machines. These data may be used for packaged chillers, including their drive units. For built-up assemblies, these data should be used for the centrifugal compressor only.

(d) *Absorption Machines.* The sound levels of absorption machines are generally masked by other noises in a mechanical equipment room. The machine usually includes one or two small pumps; steam flow noise or steam valve noise may also be present.

(e) *Built-up Refrigeration Machines.* The noise of packaged chillers, as presented in the preceding paragraphs, includes the noise of both the compressor and the drive unit. If a refrigeration system is built up of separate pieces, then the noise level estimate should include the noise of each component making up the assembly. Compressor noise levels should be taken from the packaged chiller data. Sound level data for the drive units (motors, gears, steam turbines) should be taken from the appropriate tables in this chapter or obtained from the manufacturers.

2.2 Boilers

The wide variety of blower assemblies, air and fuel inlet arrangements, burners, and combustion chambers provides such variability in the noise data that it is impossible to correlate noise with heating capacity. Noise levels at the normalized 0.9-m distance may be as high for the smallest as for the largest units. The estimated noise levels given in Table 2 (Ref. 3, p. 7-6) are believed applicable for all boilers in the size range of 50–2000 boiler horsepower (1 bhp = 33,500 Btu/h or 9818 W). These noise levels apply to the front of the boiler, so when other distances are of concern, the distance should always be

TABLE 2 Normalized Sound Pressure Levels for Boilers

Octave Frequency Band (Hz)	Sound Pressure Level (dB) 50–2000 bhp
31	90
63	90
125	90
250	87
500	84
1000	82
2000	80
4000	76
8000	70
A-weighted, dB(A)	88

taken from the front surface of the boiler. Noise levels are lower off the side and rear of the typical boiler.

2.3 Cooling Towers

The generalizations drawn here may not apply exactly to all cooling towers and condensers, but the data are useful for laying out cooling towers and their possible noise control treatments. For aid in identification, four general types of cooling towers are sketched in Fig. 1: centrifugal-fan blow-through type, axial-flow blow-through type (with the fan or fans located on a side wall, induced-draft propeller type, and "underflow" forced-draft propeller type (with the fan located under the assembly).

(a) *Sound Power Level of Propeller-Type Cooling Towers.* The approximate overall and A-weighted sound power levels of propeller-type cooling towers are given by (Ref. 2, p. 7-9):

$$L_w = 96 + 10\log(\text{fan kW}) \quad \text{for overall } L_w, \tag{1}$$

$$L_{wa} = 87 + 10\log(\text{fan kW}) \quad \text{for A-weighted } L_w, \tag{2}$$

where fan kW is the nameplate power rating of the motor that drives the fan. Octave band sound power levels can be obtained by subtracting the values of Table 3 from the overall L_w obtained by Eq. (1).

(b) *Sound Power Level of Centrifugal-Fan Cooling Towers.* The approximate overall and A-weighted sound power levels of centrifugal-fan cooling towers are given by (Ref. 3, p. 7-10):

$$L_w = 86 + 10\log(\text{fan kW}) \quad \text{for overall } L_w, \tag{3}$$

$$L_{wa} = 79 + 10\log(\text{fan kW}) \quad \text{for A-weighted sound power level.} \tag{4}$$

When more than one fan or cooling tower is used, the fan power rating should be the total motor drive rating of all fans or towers. Octave band sound power levels can be obtained by subtracting the values of Table 3 from the overall L_w obtained with Eq. (3).

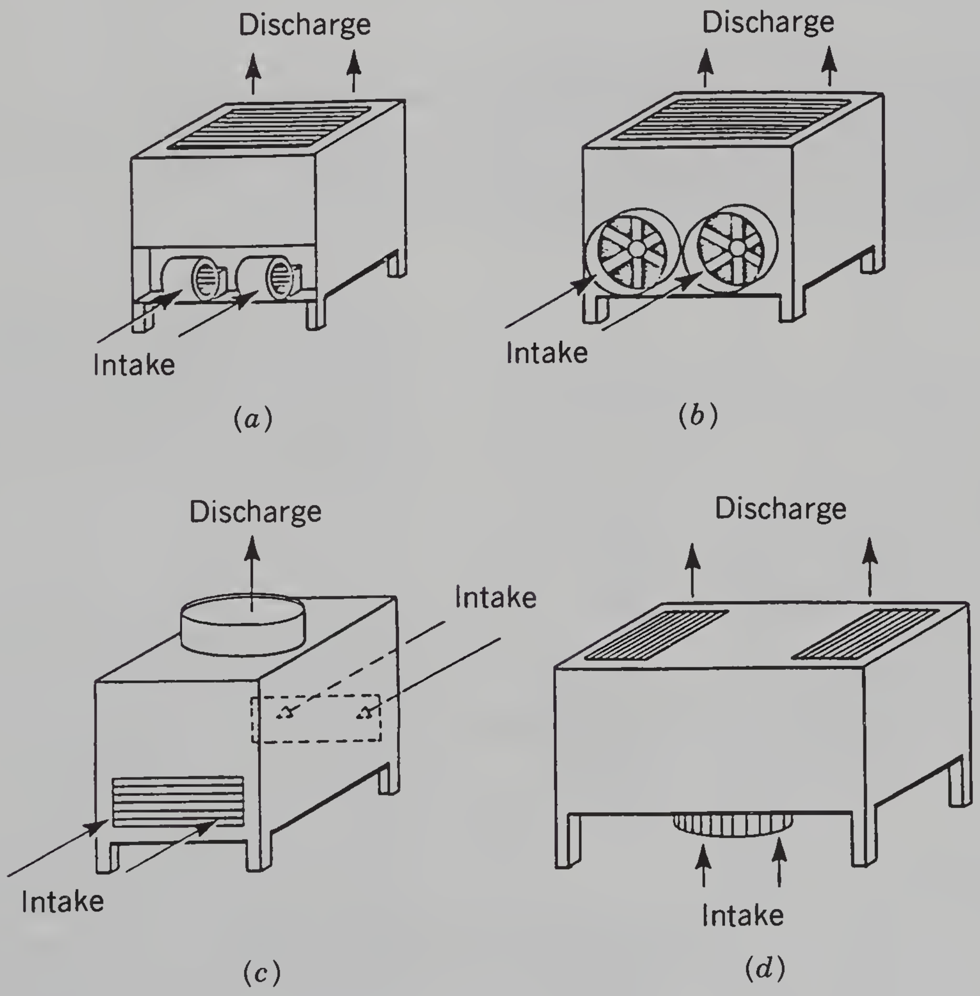


Fig. 1 Principal types of cooling towers: (*a*) centrifugal-fan blow-throw type; (*b*) axial-flow blow-through type (*c*) induced-draft propeller type; (*d*) forced-draft propeller-type "underflow."

TABLE 3 Frequency Adjustments for Cooling Towers to Obtain Octave Band PWLs

Octave Frequency Band (Hz)	Centrifugal Fan	Propeller Fan
31	6	8
63	6	5
125	8	5
250	10	8
500	11	11
1000	13	15
2000	12	18
4000	18	21
8000	25	29
A-weighted, dB(A)	7	9

(c) *Directionality of Cooling Towers.* Cooling towers usually radiate different amounts of sound in different directions, and the directional corrections of Table 4 should be made to the octave band sound power levels. These corrections apply to the five principal directions from a cooling tower, that is, in a direction perpendicular to each of the four sides and to the top of the tower. If it is necessary to estimate the sound pressure level at some direction other than the principal directions, it is common practice to interpolate between the values given for the principal directions.

(d) *Close Proximity Sound Levels of Cooling Towers.* Sound power level data usually will not give accurate calculated sound pressure levels at very close distances to large-size sources, such as cooling towers. The data of Table 5 may be used where it is required to estimate close-in sound pressure levels at nearby walls, floors, or in closely confined spaces (at 0.9–2 m distances from inlet and discharge openings).

(e) *Noise Reduction with Reduced Fan Speed.* When it is practical to do so, the cooling tower fan can be reduced to half speed in order to reduce noise. Half speed produces approximately two-thirds cooling capacity and approximately 8- to 10-dB noise reduction in the octave bands that contain most of the fan-induced noise. For half-speed operation, the octave band L_p or L_w of full-speed cooling tower noise may be reduced by the following amounts, where fB is the blade passage frequency and is calculated from the relation fB = number of fan blades × shaft rpm/60:

TABLE 4 Correction to Average PWLs for Directional Effects of Cooling Towers

Cooling Tower Type	31 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Centrifugal-Fan Blow-Through Type									
Front (fan inlet)	+3	+3	+2	+3	+4	+3	+2	+2	+2
Side (enclosed)	0	0	0	−2	−3	−4	−5	−5	−5
Rear (enclosed)	0	0	−1	−2	−3	−4	−5	−6	−6
Top (discharge)	−3	−3	−2	0	+1	+2	+3	+4	+5
Axial-Flow Blow-Through Type									
Front (fan inlet)	+2	+2	+4	+6	+6	+5	+5	+5	+5
Side (enclosed)	+1	+1	+1	−2	−5	−5	−5	−5	−4
Rear (enclosed)	−3	−3	−4	−7	−7	−7	−8	−11	−8
Top (discharge)	−5	−5	−5	−5	−2	0	0	+2	+1
Induced-Draft Propeller-Type									
Front (air inlet)	0	0	0	+1	+2	+2	+2	+3	+3
Side (enclosed)	−3	−3	−3	−3	−3	−3	−4	−5	−6
Top (discharge)	+3	+3	+3	+3	+3	+4	+4	+3	+3
"Underflow" Forced-Draft Propeller Type									
Any side	−1	−1	−1	−2	−2	−3	−3	−4	−4
Top	+2	+2	+2	+3	+3	+4	+4	+5	+5

TABLE 5 Approximate Close-in SPLs near the Intake and Discharge Openings of Various Cooling Towers (0.9–2 m distance)

Octave Band	31 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Centrifugal-Fan Blow-Through Type									
Intake	85	85	85	83	81	79	76	73	68
Discharge	80	80	80	79	78	77	76	75	74
Axial-Flow Blow-Through Type (including "underflow" type)									
Intake	97	100	98	95	91	86	81	76	71
Discharge	88	88	88	86	84	82	80	78	76
Propeller-Fan Induced-Draft Type									
Intake	94	96	94	92	88	83	78	72	76
Discharge	99	103	99	95	90	85	81	76	70

Octave Band That contains	Noise Reduction Due to Half Speed
$\frac{1}{8}$ fB	3 dB
$\frac{1}{4}$ fB	6 dB
$\frac{1}{2}$ fB	9 dB
1 fB	9 dB
2 fB	9 dB
4 fB	6 dB
8 fB	3 dB

If the blade passage frequency is not known, it may be assumed to fall in the 63-Hz band for propeller-type cooling towers and in the 250-Hz band for centrifugal cooling towers. Waterfall noise usually dominates in the upper octave bands, and it would not change significantly with reduced fan speed.

(f) *Close Confined Locations.* Most of the preceding discussion assumes that cooling towers will be used in outdoor locations. If they are located inside enclosed mechanical equipment rooms or within courts formed by several solid walls, the sound patterns will be distorted. In such instances, the L_w of the tower (or appropriate portions of the total L_w) can be placed in that setting, and the enclosed or partially enclosed space can be likened to a room having certain estimated amounts of reflecting and absorbing surfaces. In the absence of a detailed analysis of cooling tower noise levels inside enclosed spaces, it is suggested that the close-in noise levels of Table 5 be used as approximations.

2.4 Pumps

The overall and A-weighted normalized sound pressure levels for pumps are given in Table 6 (Ref. 3, p. 7-16) and are representative of most pumps encountered in commercial situations. The pump power rating is taken as the nameplate power of the drive motor. For pump ratings under 75 kW, the radiated noise increases with 10 log of the pump rating, but above 75 kW the noise changes more slowly with increasing power (i.e., 3 log of the pump rating). Octave band sound pressure lev-

TABLE 6 Overall and A-Weighted Sound Pressure Levels for Pumps

Speed Range (rpm)	Drive Motor Nameplate Power	
	Under 75 kW	Above 75 kW
Overall Sound Pressure Level (dB)		
3000–3600	72 + 10 log kW	85 + 3 log kW
1600–1800	75 + 10 log kW	88 + 3 log kW
1000–1500	70 + 10 log kW	83 + 3 log kW
450–900	68 + 10 log kW	81 + 3 log kW
A-Weighted Sound Level [dB(A)]		
3000–3600	70 + 10 log kW	82 + 3 log kW
1600–1800	73 + 10 log kW	86 + 3 log kW
1000–1500	68 + 10 log kW	81 + 3 log kW
450–900	66 + 10 log kW	79 + 3 log kW

TABLE 7 Octave Band Frequency Adjustments for Pumps

Octave Frequency Band (Hz)	Value Substracted from Overall SPL (dB)
31	13
63	12
125	11
250	9
500	9
1000	6
2000	9
4000	13
8000	19
A-weighted, dB(A)	2

els are obtained by subtracting the values of Table 7 from the overall sound pressure level of Table 6. Pumps intended for high-pressure operation have smaller clearances between the blade tips and the cutoff edge and, as a result, may have higher noise peaks in the octave bands containing the impeller blade passage frequency and its first harmonic. These would usually fall in the 1000- and 2000-Hz octave bands.

2.5 Fans

Manufacturers often furnish in-duct sound power level data of their fans on request. However, in early designs the specific model or manufacturer is not known. In these cases the octave band sound power of a fan can be estimated by[4]

$$L_w = \mathrm{Kw} + 10 \log Q + 20 \log P + \mathrm{BFI} + C, \qquad (5)$$

where L_w is the total in-duct sound power level of the fan (inlet and outlet), Kw is the specific sound power level for the particular fan design, Q is the volume flow rate in cubic metres per second, and P is the static pressure produced by the fan (in kilopascals). Values of Kw for the octave bands and for various basic fan blade designs are given in Part A of Table 8. The blade passage frequency of the fan is obtained from

$$\frac{\text{Fan rpm} \times \text{number of blades}}{60}, \qquad (6)$$

and the *blade frequency increment* (BFI) (in decibels) is added to the octave band sound power level in the octave in which the blade passage frequency occurs. It is best to obtain the number of blades and the fan rotational speed from the manufacturer to calculate the blade passage frequency. In the event this information is not available, Part B of Table 8 provides the usual blade passage frequency. The estimates given by this method assume ideal inlet and outlet flow conditions and operation of the fan at its peak design condition. The noise is quite critical to these conditions and increases significantly for deviations from ideal. Part C of Table 8 provides a correction factor for off-peak fan operation. The operating static efficiency of a fan may be calculated by

$$\text{Static efficiency} = \frac{Q \times P}{1002 \times \mathrm{BkW}} \qquad (7)$$

where Q is air volume in liters per second, P is total static pressure in kilopascals, and BkW is the brake power in kilowatts. The ratio of calculated static efficiency to peak efficiency determines the correction factor C. For cursory design the maximum peak efficiency can be taken as 65% for centrifugal fans with forward curved blades, 80% for centrifugal fans with backward inclined blades, and 70% for vaneaxial fans.

The fan housing and its nearby connected ductwork radiate fan noise into the fan room. Because of the number of variables involved, there exists no simple analysis procedure for estimating the L_w of the noise radiated by the housing and ductwork. However, Table 9 offers a method to estimate of this type of noise. These are simply deductions, in decibels, from the induct fan noise.

2.6 Air Compressors

Two types of air compressors are frequently found in buildings: One is a relatively small compressor, usually under 7 kW, used to provide a high-pressure air supply for operating the controls of the ventilation system, and the other is a medium-size compressor, possibly up to 75 kW, used to provide "shop air" to maintenance shops, machine shops, and laboratory spaces, or to provide ventilation system control pressure for large buildings. Normalized sound pressure levels for air compressors are given in Table 10 (Ref. 3, p. 7-19).

2.7 Emergency Generators

Emergency generators are typically driven by reciprocating engines or small gas turbines. As a general rule the sound level of the generator is not significant with respect to the driver:

TABLE 8 Specific Sound Power Level Kw, Blade Frequency Increment, and Off-Peak Correction for Use in Eq. (5)

Part A. Specific Sound Power Levels for Inlet or Outlets and Blade Frequency Increments (BFI) for Fans[a]

		Octave Band Center Frequency (Hz)								
Fan and Blade Type	Size and Operation	63	125	250	500	1000	2000	4000	8000	BFI
Centrifugal										
Airfoil, Backward	Over 750 mm dia.	85	85	84	79	75	68	64	62	3
Curved and Inclined	Under 750 mm dia.	90	90	88	84	79	73	69	65	3
Centrifugal										
Forward Curved	All sizes	98	98	88	81	81	76	71	66	2
Centrifugal										
Radial over 1,000 mm	Low pressure, SP 1–2.5 kPa	101	92	88	84	82	77	74	71	7
	Mid pressure, SP 2.5–5 kPa	103	99	90	87	83	78	74	71	8
	High pressure, SP 5–15 kPa	106	103	98	93	91	89	86	83	8
Vaneaxial	Hub ratio 0.3–0.4	94	88	88	93	92	90	83	79	6
	Hub ratio 0.4–0.6	94	88	81	88	86	81	75	73	6
	Hub ratio 0.6–0.8	98	97	96	96	94	92	88	85	6
Tubeaxial	Over 1000 mm dia.	96	91	92	94	92	91	84	82	7
	Under 1000 mm dia.	93	92	94	98	97	96	88	85	7
Propeller	All sizes	93	96	103	131	100	97	91	87	5

Part B. Octave Band in Which Blade Frequency Increment (BFI) Occurs[b]

Fan Type	Octave Band in Which BFI Occurs (Hz)
Centrifugal	
Airfoil, backward curved and backward inclined	250
Forward curved	500
Radial blade	125
Vaneaxial	125
Tubeaxial	63
Propeller	63

Part C. Correction Factor C for Off-peak Operation

Static Efficiency % of Peak	Correction Factor (dB)
90–100	0
85–89	3
75–84	6
65–74	9
55–64	12
50–54	15

[a] Kw for total subtract 3 dB for inlet or outlet level.

[b] Use for estimating purposes only. Where actual fan is known, use manufactures' data.

(a) *Reciprocating Engines.* Reciprocating engines have three sound sources of interest; the engine casing, the air inlet into the engine, and the exhaust from the engine. The L_w of the noise radiated by the casing of a natural-gas or diesel reciprocating engine is given by (Ref. 3, p. 7-20)

$$L_w = 94 + 10\log(\text{rated kW}) + A + B + C + D, \qquad (8)$$

where L_w is the overall sound power level (in decibels);

TABLE 9 Octave Band Adjustments for Estimating the PWL of Noise Radiated by a Fan Housing and Its Nearby Connected Duct Work

Octave Frequency Band (Hz)	Value Subtracted from In-Duct Fan Noise (dB)
31	—
63	0
125	0
250	5
500	10
1000	15
2000	20
4000	22
8000	25

TABLE 10 Normalized Sound Pressure Levels for Air Compressors

Octave Frequency Band (Hz)	Air Compressor Power Range		
	0.7–1.5 kW	2–7 kW	8–55 kW
31	82	87	92
63	81	84	87
125	81	84	87
250	80	83	86
500	83	86	89
1000	86	89	92
2000	86	89	92
4000	84	87	90
8000	81	84	87
A-weighted, dB(A)	91	94	97

TABLE 11 Correction Terms Applied to Eq. (8) for Estimating the Overall PWL of the Casing Noise of a Reciprocating Engine

Speed correction term *A*
- Under 600 rpm, −5 dB
- 600–1500 rpm, −2 dB
- Above 1500 rpm, 0 dB

Fuel correction term *B*
- Diesel fuel only, 0 dB
- Diesel and/or natural gas, 0 dB
- Natural gas only (may have small amount of "pilot oil"), −3 dB

Cylinder arrangement term *C*
- In-line, 0 dB
- V-type, −1 dB
- Radial, −1 dB

Air intake correction term *D*
- Unducted air inlet to unmuffled roots blower, +3 dB
- Ducted air from outside the room or into muffled roots blower, 0 dB
- All other inlets to engine (with or without turbochargers), 0 dB

TABLE 12 Frequency Adjustments for Casing Noise of Reciprocating Engines

	Value Subtracted from Overall PWL (dB)			
		Engine Speed 600–1500 rpm		
Octave Frequency Band (Hz)	Engine Speed under 600 rpm	Without Roots Blower	With Roots Blower	Engine Speed over 1500 rpm
31	12	14	22	22
63	12	9	16	14
125	6	7	18	7
250	5	8	14	7
500	7	7	3	8
1000	9	7	4	6
2000	12	9	10	7
4000	18	13	15	13
8000	28	19	26	20
A-weighted, dB(A)	4	3	1	2

"rated kW" is the engine manufacturer's continuous full-load rating for the engine, and *A*, *B*, *C*, and *D* are correction terms given in Table 11. Octave band sound power levels can be obtained by subtracting the Table 12 values from the overall L_w given by Eq. (8). The octave band corrections are different for the different engine speed groups. For small engines, under about 335 kW, the air intake noise is usually radiated close to the engine casing, so it is not easy or necessary to separate these two sources; and the engine casing noise may be considered as including air intake noise (from both naturally aspirated and turbocharged engines).

Most large engines have turbocharges at their inlet to provide pressurized air into the engine for increased performance. Turbine configuration and noise output can vary appreciably, but an approximation of the L_w of the turbocharged inlet noise is given by

$$L_w = 95 + 5\log(\text{rated kW}) - \frac{L}{1.8}, \tag{9}$$

where L_w and rated kW are as previously defined and L is the length, in metres, of a ducted inlet to the tur-

TABLE 13 Frequency Adjustments for Turbocharger Air Inlet Noise and Engine Exhaust

Octave Frequency Band (Hz)	Turbocharger Air Intake	Unmuffled Exhaust
31	4	5
63	11	9
125	13	3
250	13	7
500	12	15
1000	9	19
2000	8	25
4000	9	35
8000	17	43
A-weighted, dB(A)	3	12

bocharger. For many large engines, the air inlet may be ducted to the engine from a fresh air supply or a location outside the room or building. The octave band values given in Table 13, for air inlet, are subtracted from the overall L_w of Eq. (9) to obtain the octave band sound power levels of turbocharged inlet noise.

The sound power level of the noise radiated from the unmuffled exhaust of an engine is given by

$$L_w = 120 + 10\log(\text{rated kW}) - T - \frac{L}{1.2}, \qquad (10)$$

where T is the turbocharger correction term ($T = 0$ dB for an engine without a turbocharger and $T = 6$ dB for an engine with a turbocharger) and L is the length, in metres, of the exhaust pipe. The octave band sound power levels of unmuffled exhaust noise are obtained by subtracting the values of Table 13, for exhaust, from the overall L_w derived from Eq. (10). If the engine is equipped with an exhaust muffler, the final noise radiated from the end of the tailpipe is the L_w of the unmuffled exhaust minus the insertion loss, in octave bands, of the muffler.

(b) *Small Gas Turbine.* As with reciprocating engines, the three principal sound sources of turbine engines are the engine casing, the air inlet, and the exhaust. Most gas turbine manufacturers will provide sound power estimates of these sources. However, when these are unavailable, the overall sound power levels of these three sources, with no noise reduction treatments, are as given in the following equations (Ref. 3, p. 7-24):

$$L_w = 120 + 5\log(\text{rated MW}) \quad \text{for engine casing noise,} \qquad (11)$$

TABLE 14 Frequency Adjustments (in dB) for Gas Turbine Engine Noise Sources

Octave Frequency Band (Hz)	Value Subtracted from Overall PWL (dB)		
	Casing	Inlet	Exhaust
31	10	19	12
63	7	18	8
125	5	17	6
250	4	17	6
500	4	14	7
1000	4	8	9
2000	4	3	11
4000	4	3	15
8000	4	6	21
A-weighted, dB(A)	2	0	4

$$L_w = 127 + 15\log(\text{rated MW}) \quad \text{for air inlet noise,} \qquad (12)$$

$$L_w = 133 + 10\log(\text{rated MW}) \quad \text{for exhaust noise,} \qquad (13)$$

where rated MW is the maximum continuous full-load rating of the engine in megawatts.

For casing and inlet noise, particularly strong high-frequency sounds may occur at several of the upper octave bands. However, which bands contain the tones will depend on the specific design of the turbine and, as such, will differ between models and manufacturers. Therefore, the octave band adjustments of Table 14 allow for possible peaks in several different bands. Because of this randomness of peak frequencies, the A-weighted levels may also vary from the values quoted.

The turbine manufacturer often provides the turbine casing with a protective thermal wrapping or an enclosing cabinet, either of which can give some noise reduction. Table 15 suggests the approximate noise reduction for casing noise that can be assigned to different types of engine enclosures. Refer to the notes of Table 15 for a broad description of the enclosures. The values of Table 15 may be subtracted from the octave band sound power levels of casing noise to obtain the adjusted sound power levels of the covered or enclosed casing. An enclosure specifically designed to control casing noise may provide larger noise reduction values than those in the table. However, it should be noted that the performance of enclosures that are supported on the same structure as the gas turbine will be limited by structure-borne sound. For this reason care should be used in applying laboratory data of enclosure performance to the estimation of sound reduction of gas turbine enclosures.

The directivity of the exhaust stack should be considered in the assessment of noise control. Table 16

TABLE 15 Approximate Noise Reduction of Gas Turbine Engine Casing Enclosures

Octave Frequency Band (Hz)	Noise Reduction (dB)				
	Type 1[a]	Type 2[b]	Type 3[c]	Type 4[d]	Type 5[e]
31	2	4	1	3	6
63	2	5	1	4	7
125	2	5	1	4	8
250	3	6	2	5	9
500	3	6	2	6	10
1000	3	7	2	7	11
2000	4	8	2	8	12
4000	5	9	3	8	13
8000	6	10	3	8	14

[a] Glass fiber or mineral wool thermal insulation with lightweight foil cover over the insulation.

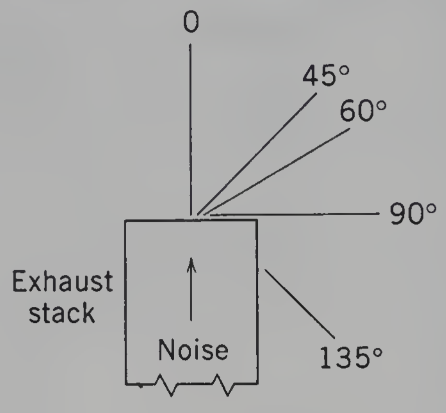
[b] Glass fiber or mineral wool thermal insulation with minimum 20 gage aluminum or 24 gauge steel or 12-mm-thick plaster cover over the insulation.

[c] Enclosing metal cabinet for the entire packaged assembly, with *open* ventilation holes and with *no* acoustical absorption lining inside the cabinet.

[d] Enclosing metal cabinet for the entire packaged assembly, with *open* ventilation holes and *with* acoustical absorption lining inside the cabinet.

[e] Enclosing metal cabinet for the entire packaged assembly, with all ventilation holes into the cabinet muffled and with acoustical absorption lining inside the cabinet.

TABLE 16 Approximate Directivity Effect of a Large Exhaust Stack Compared to a Nondirectional Source of the Same Power



Octave Frequency Band (Hz)	Relative Sound Level for Indicated Angle from Axis				
	0°	45°	60°	90°[a]	135° and Larger[a]
31	8	5	2	−2	−3
63	8	5	2	−3	−4
125	8	5	2	−4	−6
250	8	6	2	−6	−8
500	9	6	2	−8	−10
1000	9	6	2	−10	−13
2000	10	7	0	−12	−16
4000	10	7	−1	−14	−18
8000	10	7	−2	−16	−20

[a] For air intake openings subtract 3 dB from the values in the 90° and 135° columns, i.e., −2 − 3 = −5 dB for 31 cps at 90°.

gives the approximate directivity effect of a large exhaust opening. This can be used for either a horizontal or vertical stack exhausting hot gases. These values also apply for a large-area intake opening into a gas turbine for the 0°–60° range; for the 90°–135° range, subtract an additional 3 dB from the already negative-valued quantities. For small openings in a wall, such as for ducted connections to a fan intake or discharge, use approximately one-half the directivity effect of Table 16 (as applied to intake openings) for the 0°–90° region. For angles beyond 90° estimate the effect of the wall as a barrier.

2.8 Electric Motors

Electric motors used in commercial buildings are usually either a total enclosed fan-cooled (TEFC) motor or an open drop-proof motor (DRPR):

(a) *TEFC Motors.* The normalized unweighted sound pressure level for TEFC motors follow approximately the following relationships (Ref. 3, p. 7-28):

$$L_p = 17 + 17\log\ \text{kW} + 15\log\ \text{rpm} \quad \text{for power ratings under 37 kW,} \tag{14}$$

$$L_p = 28 + 10\log\ \text{kW} + 15\log\ \text{rpm} \quad \text{for power ratings above 37 kW,} \tag{15}$$

where kW is the nameplate motor range and rpm is the motor shaft speed. For motors above 300 kW, the calculated noise value for a 300-kW motor should be used. Equation (15) is not applicable to large commercial motors in the power range of 750–4000 kW. The octave band corrections for TEFC motors are given in Table 17. Some TEFC motors produce strong tonal sounds in the 500-, 1000-, or 2000-Hz octave bands because of the cooling fan blade frequency. The octave band frequency adjustments of Table 17 allow for a moderate amount of these tones, but a small percentage of motors may still exceed these calculated levels by as much as 5–8 dB. When specified, motors that are quieter than these calculated values by 5–10 dB can be purchased.

(b) *DRPR Motors.* The normalized unweighted sound pressure level for DRPR motors follow approximately the relationships (Ref. 3, p. 7-29):

TABLE 17 Frequency Adjustments for TEFC and DRPR Electric Motors

Octave Frequency Band (Hz)	TEFC	DRPR
31	14	9
63	14	9
125	11	7
250	9	7
500	6	6
1000	6	9
2000	7	12
4000	12	18
8000	20	27
A-weighted, dB(A)	1	4

$$L_p = 12 + 17\log\ \text{kW} + 15\log\ \text{rpm} \quad \text{for power ratings under 37 kW}, \tag{16}$$

$$L_p = 23 + 10\log\ \text{kW} + 15\log\ \text{rpm} \quad \text{for power ratings above 37 kW}. \tag{17}$$

For motors above 300 kW, the calculated noise value for a 300-kW motor should be used. The octave band corrections for DRPR motors are given in Table 17.

2.9 Steam Turbines

Steam turbines are sometimes used as primary or backup drivers for chillers, pumps, and air compressors. The noise levels of steam turbines are found generally to increase with increasing power rating, but it has not been possible to attribute any specific noise characteristics with speed or turbine blade passage frequency. Suggested normalized sound pressure levels for steam turbines, with a power range of 370–11,000 kW, are given in Table 18 (Ref. 3, p. 7-31).

2.10 Gears

Speed reduction or increasing gear units are used to transfer power between the primary driver and driven equipment, when there is a speed difference between the two. It is generally true that the noise of gears increase with increasing speed and power, but it is not possible to predict in which frequency band the gear tooth contact rate or the "ringing frequencies" will occur for any unknown gear. The possibility that these frequency components may occur in any of the upper octave bands is covered by Eq. (18), which gives the normalized octave band sound pressure level estimate for all bands at and about 125 Hz (Ref. 3, p. 7-31):

$$L_p = 79 + 3\log(\text{rpm}) + 4\log(\text{kW}), \tag{18}$$

where rpm is the speed of the slower gear shaft and kW is the horsepower rating of the gear or the power transmitted through the gear. For the 63-Hz band, 3 dB is deducted; and for the 31-Hz band, 6 dB is deducted from the Eq. (18) value.

2.11 Generators

The noise of generators, in general, can be quite variable, depending on speed, the presence or absence of air cooling vanes, clearances of various rotor parts, and so on,

TABLE 18 Normalized Sound Pressure Levels for Steam Turbines

Octave Frequency Band (Hz)	Steam Turbine Power Range		
	370–1100 kW	1100–3700 kW	3700–11,000 kW
31	86	88	90
63	91	93	95
125	91	93	95
250	88	90	92
500	85	87	89
1000	85	88	91
2000	87	91	95
4000	84	88	92
8000	76	81	86
A-weighted, dB(A)	92	95	99

TABLE 19 Approximate Overall PWL of Generators, Excluding Noise of Driver Unit

	Electrical Power Rating of Generator							
Generator Speed (rpm)	0.2 MW (dB)	0.5 MW (dB)	1 MW (dB)	2 MW (dB)	5 MW (dB)	10 MW (dB	20 MW (dB)	50 MW (dB)
600	95	99	102	105	109	112	115	119
1200	97	101	104	107	111	114	117	121
1800	98	102	105	108	112	115	118	122
2400	99	103	106	109	113	116	119	123
3600	100	104	107	110	114	117	120	124
4800	101	105	108	111	115	118	121	125

but, most of all, on the driver mechanism. Table 19 gives a reasonable approximation of the overall sound power level. It is to be noted that the L_w of the generator is usually less than that of the drive gear and less than that of the untreated engine casing. Octave band corrections to the overall L_w are given in Table 20.

2.12 Transformers

Transformer manufacturers commonly provide an average A-weighted sound level for their products. Typically this is an average of the sound levels, on a reference sound producing surface space at a distance of 0.3 m from the outline of the transformer. On the basis of field studies of many transformer installations, the sound power level in octave bands has been related to the average A-weighted sound level and the area of the four side walls of the unit. This relationship is expressed by

$$L_w = \text{average } L_{pa} + 10 \log A + C + 10, \qquad (19)$$

where A is the total surface of the four side walls of the transformer in cubic metres and C is an octave band correction that has different values for different uses, as shown in Table 21 (Ref. 3, p. 7-34). If the exact dimensions of the transformer are not known, an approximation will suffice. If in doubt, the area should be estimated on the high side. An error of 25% in area will produce a change of 1 dB in the sound power level. The most nearly applicable C value from Table 21 should be used. The C_1 value assumes normal radiation of sound. The C_2

TABLE 20 Frequency Adjustments for Generators, Without Drive Unit

Octave Frequency Band (Hz)	Value Subtracted from Overall SPL (dB)
31	11
63	8
125	7
250	7
500	7
1000	9
2000	11
4000	14
8000	19
A-weighted, dB(A)	4

TABLE 21 Octave Band Corrections Used in Eq. (20) for Obtaining PWL of Transformers in Different Installation Conditions

Octave Frequency Band (Hz)	Octave Band Corrections (dB)		
	C_1[a]	C_2[b]	C_3[c]
31	−1	−1	−1
63	+5	+8	+8
125	+7	+13	+13
250	+2	+8	+12
500	+2	+8	+12
1000	−4	−1	+6
2000	−9	−9	+1
4000	−14	−14	−4
8000	−21	−21	−11

[a] Use C_1 for outdoor location or for indoor location in a large mechanical equipment room (over about 5000 ft^3) containing many other pieces of mechanical equipment that serve as obstacles to diffuse sound and break up standing waves.

[b] Use C_2 for indoor locations in transformer vaults or small rooms (under about 5000 ft^3) with parallel walls and relatively few other large-size obstacles that can diffuse sound and break up standing waves.

[c] Use C_3 for any location where a serious noise problem would result if the transformer should become noisy above its NEMA rating, following its installation and initial period of use.

value should be used in regular-shaped confined spaces where standing waves will likely occur, which typically may produce 6 dB higher sound levels at the transformer harmonic frequencies of 120, 240, 360, 480, and 600 Hz (for 60 Hz line frequency; or other sound frequencies for other line frequencies). The C_3 value is an approximation of the noise of a transformer that has grown noisier (by about 10 dB) during its lifetime. This happens occasionally when the laminations or tie bolts become loose, and the transformer begins to buzz or rattle. In a highly critical location, it would be wise to use this value. Field measurements have shown that transformers may actually have A-weighted sound levels that range from a few decibels (2 or 3 dB) above to as much as 5 or 6 dB below the quoted A-weighted sound level. Quieted transformers that contain various forms of noise control treatments can be purchased at as much as 15–20 dB below normal A-weighted ratings.

3 PROPAGATION OF AIRBORNE NOISE FROM MECHANICAL EQUIPMENT ROOMS

The analysis of the impact of mechanical equipment, on surrounding spaces is relatively straightforward. Once the sound pressure level within the mechanical equipment room has been established, the degree of transmission to adjacent spaces can be determined with knowledge of the transmission loss properties of the walls, floor, and ceiling of the mechanical room (see Chapter 93) and the acoustical properties of the adjacent rooms (see Chapter 95).

(a) *Transmission Loss of Mechanical Room Partitions.* Transmission loss data for different partitions can be found in many publications. As a general rule this information is derived from laboratory measurements (e.g., ASTM E 90). However, due to measurement limitations, transmission loss data below 100 Hz is rarely reported. Large mechanical equipment will often produce significant acoustic energy below 100 Hz. Therefore there is a need to obtain transmission loss performance below 100 Hz, or alternatively, estimate the low-frequency performance (see Chapter 93). Quite often a single number rating, such as the Sound Transmission Loss (STC) (ASTM E 413-87) is provided. However, most of these single-number classifications are heavily weighted toward the 500–2000 Hz frequency range. While this range is suitable for evaluation of isolation for speech, some music, and most transportation noise sources, it is not suitable for the evaluation of mechanical equipment noise sources. There is growing usage of an alternative rating call the MTC (for mechanical or music transmission loss) [Roller 85]. The determination of the MTC is similar to the STC in that it uses the same reference curve and measured one-third-octave band transmission loss data. The determination of the MTC rating differs from the determination of the STC rating in that:

1. No deficiencies are allowed in the 125- and 160-Hz one-third-octave bands.
2. And if there are any surpluses above the STC contour in the 125- and 160-Hz one-third-octave bands the rating is increased by one-third of the sum of the surpluses.

Studies have indicated that, when the A-weighted sound level within the mechanical equipment room is less than the sum of the MTC rating of the partition and the RC rating of the background sound within the adjacent room, the intrusive noise should be acceptable. MTC ratings are useful as a cursory evaluation technique. Final selection of partition types should be based on a more complete analysis (e.g., octave or one-third-octave band analysis).

(b) *Openings in Walls.* An opening, such as a door, window, or louvered vent, in an exterior wall of a noisy room will allow noise to escape from that room and perhaps be disturbing to neighbors. The sound power of the sound that passes through the opening can be estimated from the equation.

$$L_w = L_p + 10 \log A, \qquad (20)$$

where L_p is the sound level in the room at the location of the opening and A is the area, in cubic metres, of the opening. For normal openings (windows or vents) without ducted connections to the noise source, it may be assumed that the sound radiates freely in all directions in front of the opening. For ducted connections from a sound source to an opening in the wall, the sound is somewhat "beamed" out of the opening and may be assumed to have a directivity effect of above one-half the amount given in Table 16 for air intake openings of large stacks.

4 VIBRATION ISOLATION OF MECHANICAL EQUIPMENT

If mechanical equipment is not provided with proper vibration isolation, acoustic energy will be transmitted into the supporting structure resulting in unwanted vibration and structure-borne sound. The isolator types and isolation guidelines presented in this chapter are based on experience with successful installation of mechanical equipment in commercial buildings.

4.1 Isolator Types and Transmissibility

A transmissibility curve is often used to indicate the general behavior of a vibration-isolated system. Transmissibility is roughly defined as the ratio of the force transmitted through the isolated system to the supporting structure to the driving force exerted by the piece of vibrating equipment (see Chapter 71 of Volume 1). Strict interpretation of transmissibility data and isolation efficiencies, however, must be adjusted for real-life situations. Usually, floors that have large column spacing will have larger deflections than floors of shorter column spacing.

The *static deflection* of a mount is simply the difference between the free-standing height of the uncompressed, unloaded isolator and the height of the compressed isolator under its static load. This difference is easily measured in the field or estimated from the manufacturer's catalog data. For this reason it is often most beneficial to specify the static deflection for the vibration isolation system, taking into account the type of isolator, the linearity of the isolator system and the rigidity of the structure of the equipment base and supporting structure.

Table 22 provides a suggested schedule for achieving various degrees of vibration isolation in normal construction. This table is based on the transmissibility curve (see Fig. 2 of Chapter 71 of Volume 1), but suggests operating ranges of the ratio of driving frequency to natural frequency. The terms "low," "fair," and "high" are merely word descriptors, but they are more meaningful than such terms as 95 or 98% isolation efficiency, which are clearly erroneous if they do not take into account the mass and stiffness of the floor slab. Vibration control recommendations given in this chapter are based on the application of this table. Table 23 lists the principal types of vibration isolators used in the isolation of building mechanical equipment and their general range of applications:

TABLE 22 Suggested Schedule for Estimating Relative Vibration Isolation Effectiveness of a Mounting System

Ratio of Driving Frequency of Source to Natural Frequency of Mount	Degree of Vibration Isolation
Below 1.4	Amplification
1.4–3	Negligible
3–6	Low
6–10	Fair
Above 10	High

(a) *Steel Spring Isolators.* Steel springs are used to support heavy equipment and to provide isolation for the typical low-frequency range of about 3–60 Hz (180- to 3600-rpm shaft speed). Steel springs have natural frequencies that fall in the range of about 1 Hz (for approximately 254 mm static deflection) to about 6 Hz (for approximately 6 mm static deflection). Springs can transmit high-frequency vibrational energy. For this reason springs are often used in series with a neoprene, compressed glass fiber, or cork-pad-type isolator, when

TABLE 23 General Types and Applications of Vibration Isolators

			Vibration Isolation Applications—Nonspecific			
			Noncritical Locations		Critical Locations	
Isolator Type	Typical Range of Static Deflection (mm)	Corresponding Approximate Range of Natural Frequency (Hz)	Rotary Equipment	Reciprocating Equipment	Rotary Equipment	Reciprocating Equipment
Steel spring	6–12	1–6	Yes	Yes	Yes	Special
Neoprene-in-shear, double deflection	6–12	4–6	Yes	Yes	Yes	No
Neoprene-in-shear, single deflection	3–6	6–10	Yes	Yes	Yes	No
Compressed block of glass fiber, 2-in. thick	0.5–4	8–20	Yes	No	Yes	No
Neoprene pad, ribbed or waffle pattern, 1–4 layers	0.15–6	6–20	Yes	No	Yes	No
Felt or cork pads or strips	0.07–2.5	1–30	See text for applications and limitations			
Air Spring	—	1–10	See text for applications and limitations			

used to support equipment directly over critical locations in a building. Unhoused "stable" steel springs are preferred over housed unstable or stable springs. Stable steel springs have a diameter that is about 0.8–1.2 times their compressed height. They have a horizontal stiffness that is approximately equal to their vertical stiffness; therefore, they do not have a tendency to tilt sideways when a vertical load is applied.

(b) *Neoprene-in-Shear Isolators.* Neoprene is a long-lasting material that, when properly shaped, can provide good vibration isolation for the conditions shown in Table 23. The mount usually has an interior hollow space that is conically shaped. The total effect of the shaping is that for almost any direction of applied load, there is a shearing action on the cross section of neoprene. A solid block of neoprene in compression is not as effective as an isolator. Manufacturers' catalogs will show the upper limit of load-handling capability of neoprene-in-shear mounts. Two neoprene-in-shear mounts are sometimes constructed in series in the same supporting bracket to provide additional static deflection. This is the double deflection mount referred to in Table 23.

(c) *Compressed Glass Fiber.* Blocks of compressed glass fiber serve as vibration isolators when properly loaded. The manufacturers have several different densities available for a range of loading conditions. Typically, a block is about 50 mm thick and has an area of about 60–130 cm^2, but other dimensions are available. These blocks attenuate high-frequency structure-borne noise, and they are often used alone, at various spacings, to support floating concrete floor slabs. The manufacturer's data should be used to determine the density and area of a block required to achieve the desired static deflection. Unless otherwise indicated, a static deflection of about 5–10% of the uncompressed height is normal. With longtime use, the material may compress an additional 5–10% of its height. This gradual change in height must be kept in mind during the designing of floating floors to meet floor lines of structural slabs.

(d) *Ribbed Neoprene Pads.* Neoprene pads with ribbed or waffle pattern surfaces are effective as high-frequency isolators. In stacks of 2–4 thickness, with each layer separated by a metal plate, they are also used for vibration isolation of low-power rotary equipment. The pads are usually about 6–10 mm thick, and they compress by about 20% of their height when loaded at about 140–350 kPa. Higher durometer pads may be loaded up to about 700 kPa. The pads are effective as isolators because the ribs provide some shearing action, and the spaces between the ribs allow lateral expansion as an axial load is applied. The manufacturer's literature should be used for proper selection of the material (load–deflection curves, durometer, surface area, height, etc.).

(e) *Felt Pads.* Felt strips or pads are effective for reducing structure-borne sound transmission in the mounting of piping and vibrating conduit. One or more layers of 3- or 6-mm-thick strips should be wrapped around the pipe under the pipe clamps that attach the piping to building structures. Felt pads will compress under long duration and high load application and should not be used alone to vibration isolate heavy equipment.

(f) *Air Springs.* Air springs are the only practical vibration isolator for very low frequencies, down to about 1 Hz or even lower for special applications. An air mount consists of pressurized air enclosed in a resilient reinforced neoprene chamber. Since the air chamber is subject to slow leakage, a system of air mounts usually includes a pressure sensing system and external air supply. A group of air mounts can be arranged to maintain very precise leveling of a base by automatic adjustment of the pressure in the individual mounts. Specific design and operation data should be obtained from the manufacturer.

4.2 Mounting Assembly Types

In this section, five basic mounting systems are described for the vibration isolation of equipment. These mounting systems are applied to specific types of equipment in Section 4.3.

(a) *Type I Mounting Assembly.* The specified equipment should be mounted rigidly on a concrete inertia block. The length and the width of the inertia block should be at least 30% greater than the length and width of the supported equipment. Mounting brackets for stable steel springs should be located off the sides of the inertia block at or near the height of the vertical center of gravity of the combined completely assembled equipment and concrete block. The clearance between the floor and the concrete inertia block shall be at least 100 mm, and provision should be allowed to check this clearance at all points under the block. The inertia block adds stability to the system and reduces motion of the system in the vicinity of the driving frequency.

(b) *Type II Mounting Assembly.* This mount is the same as the type I mount in all respects except that the mounting brackets and the top of the steel springs should be located as high as practical on the concrete inertia block but not necessarily as high as the vertical center-of-gravity position of the assembly. The steel springs can be recessed into pockets in the concrete block, but clearances around the springs should be large enough to assure no contact between any spring and any part of the mounted assembly. Provision must be made to allow positive visual inspection of the spring clearance. When this type of mounting is used for a pump, the concrete inertia block can be given a T-shape plan, and the pipes to and from the pump can be supported rigidly with the pump onto the wings of the T. In this way, the pipe elbows will not be placed under undue stress.

(c) *Type III Mounting Assembly.* The equipment or the assembly of equipment should be mounted on a steel frame that is stiff enough to allow the entire assembly to be supported on flexible point supports without fear of distortion of the frame or misalignment of the equipment. The frame should then be mounted on resilient mounts, steel springs, neoprene-in-shear mounts, or isolation pads, as the static deflection would require. If the equipment frame itself already has adequate stiffness, no additional framing is required, and the isolation mounts may be applied directly to the base of the equipment.

(d) *Type IV Mounting Assembly.* The equipment should be mounted on an array of "pad mounts." The pads may be of compressed glass fiber or of multiple layers of ribbed neoprene or waffle pattern neoprene of sufficient height and of proper stiffness to support the load while meeting the static deflection recommended in the applicable accompanying tables.

(e) *Type V Mounting Assembly (for Propeller-Type Cooling Towers).* Large, low-speed propeller-type cooling towers located on roof decks of large buildings may produce serious vibration in their buildings if adequate vibration isolation is not provided. In extreme cases, the vibration may be evident two or three floors below the cooling towers. It is recommended that the entire cooling tower assembly be isolated on stable steel springs. The springs should be in series with at least two layers of ribbed or waffle pattern neoprene if there is an acoustically critical area immediately below the cooling tower (or within about 10 m horizontally on the floor immediately under the tower). It is necessary to provide

TABLE 24 Recommended Vibration Isolation Mounting Details for Centrifugal and Axial-Flow Fans

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	Under 2	Under 600					
		600–1200				No isolation required	
		Over 1200					
	2–19	Under 600	III	—		25 mm	
		600–1200				12 mm	
		Over 1200				12 mm	
	20–150	Under 600	III	—		40 mm	
		600–1200				25 mm	
		Over 1200				12 mm	
On upper floor above noncritical area	Under 2	Under 600	III	—	25 mm	40 mm	50 mm
		600–1200	III	—	12 mm	20 mm	25 mm
		Over 1200	III	—	12 mm	12 mm	20 mm
	3–19	Under 600	II	2	25 mm	40 mm	50 mm
		600–1200	III	—	40 mm	50 mm	75 mm
		Over 1200	III	—	25 mm	40 mm	50 mm
	70–150	Under 600	II	2	50 mm	75 mm	100 mm
		600–1200	II	2	40 mm	50 mm	75 mm
		Over 1200	II	2	25 mm	40 mm	50 mm
On upper floor above critical area	Under 2	Under 600	II	2	40 mm	50 mm	75 mm
		600–1200	III	—	40 mm	50 mm	75 mm
		Over 1200	III	—	25 mm	40 mm	50 mm
	3–19	Under 600	II	3	50 mm	75 mm	100 mm
		600–1200	II	2	40 mm	50 mm	75 mm
		Over 1200	II	2	25 mm	40 mm	50 mm
	20–150	Under 600	II	3	75 mm	100 mm	125 mm
		600–1200	II	2	50 mm	65 mm	75 mm
		Over 1200	II	2	25 mm	40 mm	50 mm

limit stops on these springs to limit movement of the tower when it is emptied and to provide limited movement under wind load.

4.3 Tables of Recommended Vibration Isolation Details

Tables 24–34 (Ref. 3, pp. 9-10–9-21) provide suggested vibration isolation for the most commonly found mechanical equipment in commercial buildings. A common format is used for all the tables that summarize the recommended vibration isolation details for the various types of equipment. Addition comments are as follows:

(a) *Centrifugal and Axial-Flow Fans.* Ducts should contain flexible connections at both the inlet and discharge of the fans, and all connections to the fan assembly should be clearly flexible. The entire assembly should bounce with little restraint when one jumps up and down on the unit. Large ducts

TABLE 25 Recommended Vibration Isolation Mounting Details for Reciprocating Compressor Refrigeration Equipment Assembly (Including Motor, Gear, or Steam Turbine Drive Unit)

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	35–175	600–900	III			50 mm	
		901–1200	III			40 mm	
		1201–2400	III			25 mm	
	175–600	600–900	II	2–3		50 mm	
		901–1200	III			50 mm	
		1201–2400	III			40 mm	
On upper floor above noncritical area		600–900	II	2–3	50 mm	75 mm	100 mm
		901–1200	II	2–3	40 mm	50 mm	70 mm
		1201–2400	II	2–3	40 mm	40 mm	50 mm
		Under 600	II	3–4	75 mm	100 mm	125 mm
		600–1200	II	3–4	50 mm	75 mm	100 mm
		Over 1200	II	2–3	50 mm	50 mm	75 mm
On upper floor above critical area		600–900	II	3–4	75 mm	100 mm	125 mm
		901–1200	II	3–4	50 mm	75 mm	100 mm
		1201–2400	II	2–3	50 mm	50 mm	75 mm
		600–900	I	4–6	75 mm	100 mm	125 mm
		901–1201	II	3–5	50 mm	75 mm	100 mm
		1201–2400	II	3–4	50 mm	50 mm	75 mm

TABLE 26 Recommended Vibration Isolation Mounting Details for Rotary Screw Compressor Refrigeration Equipment Assembly (Including Motor Drive Unit)

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	350–1700	2400–4800	III			25 mm	
On upper floor above noncritical area	350–1700	2400–4800	III		25 mm	40 mm	50 mm
On upper floor above critical area	250–1700	2400–4800	II	2–3	25 mm	40 mm	50 mm

TABLE 27 Recommended Vibration Isolation Mounting Details for Centrifugal Compressor Refrigeration Equipment Assembly[a]

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	350–1750	Over 3000	III			20 mm	
	1750–1400	Over 3000	III			20 mm	
On upper floor above noncritical area	350–1750	Over 3000	III		25 mm	40 mm	50 mm
	1750–1400	Over 3000	III		40 mm	50 mm	75 mm
On upper floor above critical area	350–1750	Over 3000	II	2–3	40 mm	50 mm	75 mm
	1750–1400	Over 3000	II	3–5	40 mm	50 mm	75 mm

[a] Including condenser and chiller tanks and motor, gear, or steam turbine drive unit.

TABLE 28 Recommended Vibration Isolation Mounting Details for Absorption-Type Refrigeration Equipment Assembly

Equipment Conditions		Mounting Recommendations					
					Min. Defl. w/Col. Spacing		
Equipment Location	Cooling Capacity	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	All sizes		IV			6 mm	
On upper floor above noncritical area	All sizes		III		12 mm	20 mm	25 mm
On upper floor above critical area	All sizes		III		25 mm	40 mm	50 mm

TABLE 29 Recommended Vibration Isolation Mounting Details for Boilers

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Heating Capacity (bhp)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	Under 200		—			Not required	
	200–1000					Not required	
	Over 1000					Not required	
On upper floor above noncritical area	Under 200		III		3 mm	6 mm	12 mm
	200–1000		III		6 mm	12 mm	25 mm
	Over 1000		III		6 mm	12 mm	25 mm
On upper floor above critical area	Under 200		III		12 mm	25 mm	40 mm
	200–1000		III		25 mm	40 mm	50 mm
	Over 1000		III		25 mm	40 mm	50 mm

(cross-section area over 1.5 m^2) that are located within about 10 m of the inlet or discharge of a large fan (over 15 kW) should be supported from the floor or ceiling with resilient mounts having a static deflection of at least 6 mm.

(b) *Reciprocating-Compressor Refrigeration Equipment.* These recommendations apply also to the drive unit used with the reciprocating compressor. Pipe connections from this assembly to other equipment should contain flexible connections.

(c) *Centrifugal-Compressor Refrigeration Equipment.* The recommended vibration isolation details for this equipment include the drive unit and the condenser and evaporator tanks.

TABLE 30 Recommended Vibration Isolation Mounting Details for Propeller-Type Cooling Towers[a]

Equipment Conditions			Mounting Recommendations			
	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	Min. Defl. w/Col. Spacing	
Equipment Location					All Spacings	
On grade slab			Vibration Isolation Usually not Required			
On upper floor above noncritical area		150–300			125 mm	
	Under 20	301–600	v Install on		75 mm	Springs may be located
		Over 600			75 mm	
		150–300			150 mm	
	20–100	301–600	v Dunnage attached to		100 mm	Under drive assembly
		Over 600			75 mm	
		150–300			150 mm	
	Over 100	301–600	v Building columns only		125 mm	Or under tower base
		Over 600			100 mm	
On upper floor above critical area	Same as for location above noncritical area, except install ribbed or waffle pattern neoprene between tower and building					

[a] Where several towers are placed at the same general location, the power range for total power of all towers should be used.

TABLE 31 Recommended Vibration Isolation Mounting Details for Centrifugal-Type Cooling Towers[a]

Equipment Conditions			Mounting Recommendations				
	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	Min. Defl. w/Col. Spacing		
Equipment Location					9 m	12 m	15 m
On grade slab			Vibration isolation usually not required				
On upper floor above noncritical area		450–900			25 mm	40 mm	50 mm
	Under 20	901–1800	III		20 mm	50 mm	40 mm
		Over 1800			20 mm	50 mm	40 mm
		450–900			40 mm	50 mm	75 mm
	20–100	901–1800	III		25 mm	40 mm	50 mm
		Over 1800			20 mm	25 mm	40 mm
		450–900			50 mm	75 mm	100 mm
	Over 100	901–1800	III		40 mm	50 mm	75 mm
		Over 1800			25 mm	40 mm	50 mm
On upper floor above critical area		450–900			40 mm	100 mm	75 mm
	Under 20	901–1800	III		25 mm	40 mm	50 mm
		Over 1800			20 mm	25 mm	40 mm
		450–900			50 mm	75 mm	100 mm
	20–100	901–1800	III		40 mm	50 mm	75 mm
		Over 1800			25 mm	40 mm	50 mm
		450–900			75 mm	100 mm	150 mm
	Over 100	901–1800	III		40 mm	50 mm	75 mm
		Over 1800			25 mm	40 mm	50 mm

[a] Power is total of all fans at the same general location.

(d) *Boilers.* The recommended vibration isolation for boilers apply for boilers with integrally attached blowers. Table 24 should be followed for the support of blowers that are not directly mounted on the boiler. A flexible connection or a thermal expansion joint should be installed in the exhaust breaching between the boiler and the exhaust stack.

TABLE 32 Recommended Vibration Isolation Mounting Details for Motor–Pump Assemblies

Equipment Conditions			Mounting Recommendations				
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	Min. Defl. w/Col. Spacing		
					9 m	12 m	15 m
On grade slab	Under 15	450–900	Vibration isolation usually not required for acoustic purposes				
		901–1800					
		1801–3600					
	15–75	450–900	II	2–3		40 mm	
		901–1800	II	1.5–2.5		25 mm	
		1801–3600	II	1.5–2.5		20 mm	
	Over 75	450–900	II	2–3		50 mm	
		901–1800	II	2–3		40 mm	
		1801–3600	II	1.5–2.5		25 mm	
On upper floor above noncritical area	Under 20	450–900	II	2–3	40 mm	40 mm	75 mm
		901–1800	II	1.5–2.5	25 mm	40 mm	50 mm
		1801–3600	II	1.5–2.5	20 mm	25 mm	40 mm
	20–100	450–900	II	2–3	40 mm	50 mm	75 mm
		901–1800	II	2–3	25 mm	40 mm	50 mm
		1801–3600	II	1.5–2.5	25 mm	40 mm	50 mm
	Over 100	450–900	II	3–4	100 mm	75 mm	100 mm
		901–1800	II	2–3	40 mm	50 mm	75 mm
		1801–3600	II	2–3	25 mm	40 mm	50 mm
On upper floor above critical area	Under 20	450–900	II	3–4	40 mm	100 mm	75 mm
		901–1800	II	2–3	25 mm	75 mm	50 mm
		1801–3600	II	2–3	20 mm	50 mm	40 mm
	20–100	450–900	II	3–4	50 mm	100 mm	100 mm
		901–1800	II	2–3	40 mm	75 mm	75 mm
		1801–3600	II	2–3	20 mm	50 mm	50 mm
	Over 100	450–900	II	3–4	75 mm	100 mm	125 mm
		901–1800	II	2–3	50 mm	75 mm	100 mm
		1801–3600	II	2–3	40 mm	50 mm	75 mm

(e) *Motor-Pump Assemblies.* Electrical connections to the motors should be made with long "floppy" lengths of flexible armored cable, and piping should be resiliently supported. For most situations, a good isolation mounting of the piping will overcome the need for flexible connections in the pipe. An important function of the concrete inertia block (type II mounting) is its stabilizing effect against undue bouncing of the pump assembly at the instant of starting. This gives better long-time protection to the associated piping. These same recommendations may be applied to other motor-driven rotary devices such as centrifugal-type air compressors and motor-generator sets in the power range up to a few hundred kilowatts.

(f) *Steam Turbines.* Table 33 provides a set of general isolation recommendations for steam-turbine-driven rotary equipment, such as gears, generators, or centrifugal-type gas compressors. The material given in Table 27 applies when a steam turbine is used to drive centrifugal-compressor refrigeration equipment. The recommendations given in Table 25 apply when a steam turbine is used to drive reciprocating-compressor refrigeration equipment or reciprocating-type gas compressors.

(g) *Gears.* When a gear is involved in a drive system, vibration isolation should be provided in accordance with recommendations given for either the main power drive unit or the driven unit, whichever imposes the more stringent isolation conditions.

(h) *Transformers.* Power leads to and from the transformers should be as flexible as possible. In outdoor locations, earth-borne vibration to nearby neighbors is usually not a problem, so no vibration isolation is suggested. If vibration should become a problem, the transformer could be installed on neoprene or compressed glass fiber pads having 6 mm static deflection.

TABLE 33 Recommended Vibration Isolation Mounting Details for Steam-Turbine-Driven Rotary Equipment, Such as Gear, Generator, or Gas Compressor[a]

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	370–110	Over 3000	III			12 mm	
	1100–3700	Over 3000	III			20 mm	
	3700–11,000	Over 3000	III			25 mm	
On upper floor above noncritical area	370–1100	Over 3000	III		25 mm	40 mm	50 mm
	1100–3700	Over 3000	III		40 mm	50 mm	75 mm
	3700–11,000	Over 3000	III		50 mm	75 mm	100 mm
On upper floor above critical area	350–1100	Over 3000	II	2–3	25 mm	40 mm	50 mm
	1100–3700	Over 3000	II	2–3	40 mm	50 mm	75 mm
	3700–1100	Over 3000	II	2–3	50 mm	75 mm	100 mm

[a] Use Table 25 for reciprocating compressor driven by steam turbine; use Table 26 for centrifugal compressor driven by steam turbine.

TABLE 34 Recommended Vibration Isolation Mounting Details for Transformers

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	9 m	12 m	15 m
On grade slab	Under 10	N/A	IV	N/A		3 mm	
	10–100	N/A	IV	N/A		3 mm	
	Over 100	N/A	IV	N/A		6 mm	
On upper floor above noncritical area	Under 10	N/A	IV	N/A	3 mm	6 mm	6 mm
	10–100	N/A	III	N/A	6 mm	12 mm	12 mm
	Over 100	N/A	III	N/A	6 mm	12 mm	25 mm
On upper floor above critical area	Under 10	N/A	III	N/A	6 mm	12 mm	20 mm
	10–100	N/A	III	N/A	6 mm	20 mm	25 mm
	Over 100	N/A	III	N/A	6 mm	25 mm	40 mm

[a] N/A = Not available.

(i) *Air Compressors.* Recommended mounting details for centrifugal-type air compressors of less than 75 kW are the same as those given for motor-pump units in Table 32. The same recommendations would apply for small (under 8 kW) reciprocating-type air compressors. For reciprocating-type air compressors (with more than two cylinders) in the 8–49 kW range, the recommendations given in Table 25 apply for the particular conditions. For 8–80 kW, one or two cylinder, reciprocating-type air compressors, the recommendations of Table 35 apply. This equipment is a potentially serious source of low-frequency vibration in a building if it is not isolated. In fact, the compressor should not be located in certain parts of the building, even if it is vibration isolated. When these compressors are used, all piping should contain flexible connections, and the electrical connections should be made with flexible armored cable.

4.4 Vibration Isolation: Miscellaneous

It is good practice to isolate all piping in the mechanical equipment room that is connected to vibrating equipment with resilient ceiling hangers or from floor-mounted resilient supports. As a general rule, the first three pipe supports nearest the vibrating equipment should have a static deflection of at least one-half the static deflection of the mounting system used with that equipment. Beyond the third pipe support, the static deflection can be

TABLE 35 Recommended Vibration Isolation Mounting Details for One- or Two-Cylinder Reciprocating-Type Air Compressors

Equipment Conditions			Mounting Recommendations				
					Min. Defl. w/Col. Spacing		
Equipment Location	Power Range (kW)	Speed Range (rpm)	Mounting Type	Inertia Base Wt. Ratio	10 m	12 m	15 m
On grade slab		300–600	I	4–8		100 mm	
	Under 15	601–1200	I	2–4		50 mm	
		1201–2400	I	1–2		25 mm	
		300–600	I	6–10		125 mm	
	15–75	601–1200	I	3–6		75 mm	
		1201–2400	I	2–3		40 mm	
On upper floor above noncritical area		300–600	Not recomended (NO[a])				
	Under 15	601–1200	I	3–6	100 mm	NO[a]	NO[a]
		1201–2400	I	2–3	50 mm	4 mm	NO[a]
		300–600	Not recomended				
	15–75	601–1200	Not recomended				
		1201–2400	I	3–6	75 mm	150 mm	NO[a]
On upper floor above critical area		300–600	II	3–4	40 mm	100 mm	75 mm
	Under 15	601–1200	II	2–3	25 mm	75 mm	50 mm
		1201–2400	II	2–3	20 mm	50 mm	40 mm
	15–75	300–2400			Not Recomended		

reduced to 6 or 12 mm for the remainder of the pipe run in the mechanical equipment room. When a pipe passes through the mechanical equipment room wall, a minimum 25 mm clearance should be provided between the pipe and the hole in the wall. The pipe should be supported on either side of the hole, so that the pipe does not rest on the wall. The clearance space should then be stuffed with fibrous filler material and sealed with a nonhardening caulking compound at both wall surfaces. Vertical pipe chases through a building should not be located beside acoustically critical areas. If they are located beside critical areas, pipes should be resiliently mounted from the walls of the pipe chase for a distance of at least 3 m beyond each such area, using both low-frequency and high-frequency isolation materials. Pipes to and from the cooling tower should be resiliently supported for their full length between the cooling tower and the associated mechanical equipment room. Steam pipes should be resiliently supported for their entire length of run inside the building. Resilient mounts should have a static deflection of at least 12 mm.

Whenever a steel spring isolator is used, it should be in series with a neoprene isolator. For ceiling hangers, a neoprene washer or grommet should always be included; and if the pipe hangers are near very critical areas, the hanger should be a combination hanger that contains both a steel spring and a neoprene-in-shear mount.

During inspection, the hanger rods should be checked to ensure they are not touching the sides of the isolator housing and thereby shorting-out the spring.

To be at all effective, a flexible pipe connection should have a length that is approximately 6–10 times its diameter. Tie rods should not be used to bolt the two end flanges of a flexible connection together. Flexible connections are either of the bellows type or are made up of wire-reinforced neoprene piping, sometimes fitted with an exterior braided jacket to confine the neoprene.

When a mechanical equipment room is located directly over or near a critical area, it is usually desirable to isolate most of the "nonvibrating equipment" with a simple mount made up of one or two pads of neoprene, or a 25- or 50-mm layer of compressed glass fiber. Heat exchangers, hot water heaters, water storage tanks, large ducts, and some large pipe stands may not themselves be noise sources, yet their pipes or their connections to vibrating sources transmit small amounts of vibrational energy that they then may transmit into the floor. A simple minimum isolation pad will usually prevent this noise transfer.

REFERENCES

1. *Heating Ventilating and Air-Conditioning Systems and Equipment*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA. 1992. Chap-

ters 18 ("Fans"), 28 ("Boilers"), 35 ("Compressors"), 37 ("Cooling Towers"), 39 ("Centrifugal Pumps"), and 41 ("Engines and Turbine Drives").

2. R. Jorgensen (Ed.), *Fan Engineering*, 8th ed., Buffalo Forge Company, Buffalo, NY, 1983, Part II, Chapters 9–19.
3. "Noise and Vibration Control for Mechanical Equipment," Army Manual TM 5-805-4, Headquarters Department of Army, The Air Force and Navy, Washington, DC, 1983.
4. B. J. Graham and R. M. Hoover, "Fan Noise," in C. M. Harris (Ed.), *The Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991, pp. 41.4–41.20.

PART X

ACOUSTICAL SIGNAL PROCESSING

99

INTRODUCTION

ANDREW F. SEYBERT

1 INTRODUCTION

In the present context, acoustical signal processing is the analysis of sound or vibration data for the purpose of extracting information important for understanding the underlying physical mechanisms of noise or vibration. While the data are usually signals obtained from experiments via transducers, the data may also be obtained from computer simulations.

In this chapter and the two that follow, the basic concepts of acoustical signal processing are discussed. In Chapter 100, the statistical theory of signals for stationary and nonstationary signals is presented. Of primary importance here is the concept of the frequency spectrum, probably the single most important quantity that can be obtained from an acoustical signal. In Chapter 100 the frequency spectrum for periodic, random, and transient signals is discussed. Also presented in Chapter 100 are the time-domain (autocorrelation) function, the probability density function, the cross-correlation function, the coherence function, and the frequency response function.

In Chapter 101, considerations of the practical implementation of the acoustical signal processing are discussed. The need for this discussion is apparent when one considers that in practice one does not analyze continuous signals of infinite duration. Modern acoustical signal processing utilizes a finite amount of discrete (sampled) acoustical data from which all required information must be obtained. As one is now working with limited data, the information obtained becomes fuzzy, that is, one must tolerate a certain amount of error or uncertainty. Chapter 101 deals with the implications of practical realities such as signal truncation, sampling, windowing of the signal, random and bias errors, and confidence limits.

An important example of acoustical signal processing is presented in Chapter 102. In that chapter, the fundamentals of acoustical holography are discussed. Acoustical holography is an intensive use of acoustical signal processing techniques. Other applications of acoustical signal processing may be found in Chapter 156 (Sound Intensity).

2 AN EXAMPLE OF ACOUSTICAL SIGNAL PROCESSING

The most common type of signal in acoustics is that obtained from a measurement of sound pressure using, for example, a microphone or a hydrophone. The sound pressure fluctuation is converted to a voltage signal by the transducer and is amplified. At this point the signal may be displayed as a function of time or analyzed in any number of ways to reveal its frequency content, its amplitude distribution, and so on, depending on what information is desired.

In most applications, one is interested in the frequency content of the signal, or the *frequency spectrum*. There are two primary reasons for obtaining frequency information about a signal. First, the response of the ear and the sensation of sound in humans is strongly dependent on frequency. Second, the physical processes of sound emission, propagation, diffraction, and transmission are all frequency dependent.

In studies of sound emission, for example, it is usually very difficult to identify and quantify the contribution of individual noise sources or noise mechanisms (such as gear noise) by examining the time history of the

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

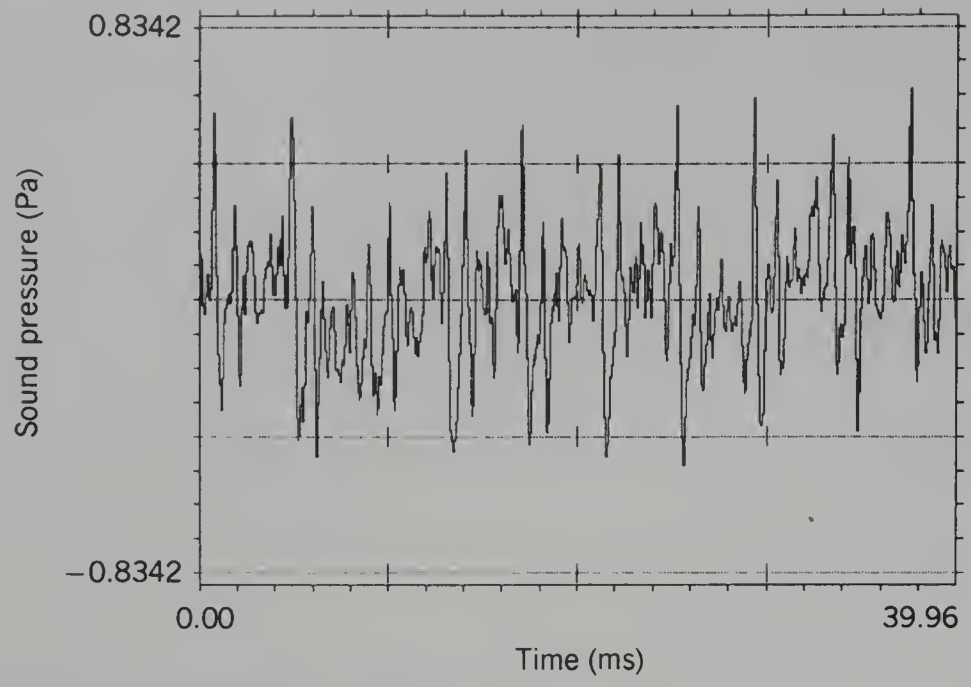


Fig. 1 Electric fan noise signal.

microphone signal. However, by analyzing the frequency spectrum of the microphone signal, a noise control engineer can identify offending components of the noise and correlate this information to underlying mechanisms of noise emission in the system under study. For example, it is well-known that most fans will produce a spectrum of noise that includes a component located at a frequency equal to the number of blades multiplied by the rotational speed of the fan in revolutions per second. This information may be used to diagnose a noise problem and assist in its solution.

Acoustic signals may be classified as either *deterministic* or *random*. Deterministic signals are those that originate from processes that either repeat (e.g., rotating machinery) or are transient (e.g., impacts or explosions). Random signals are those that originate from processes so complex that they have no apparent deterministic nature. The sound produced by a turbulent boundary layer or by high-speed airflows are typical examples of random signals.

All practical signals in acoustics are combinations of the deterministic and random signals in varying degrees. A good example is the noise produced by an electric-motor-driven fan such as found in electric appliances and

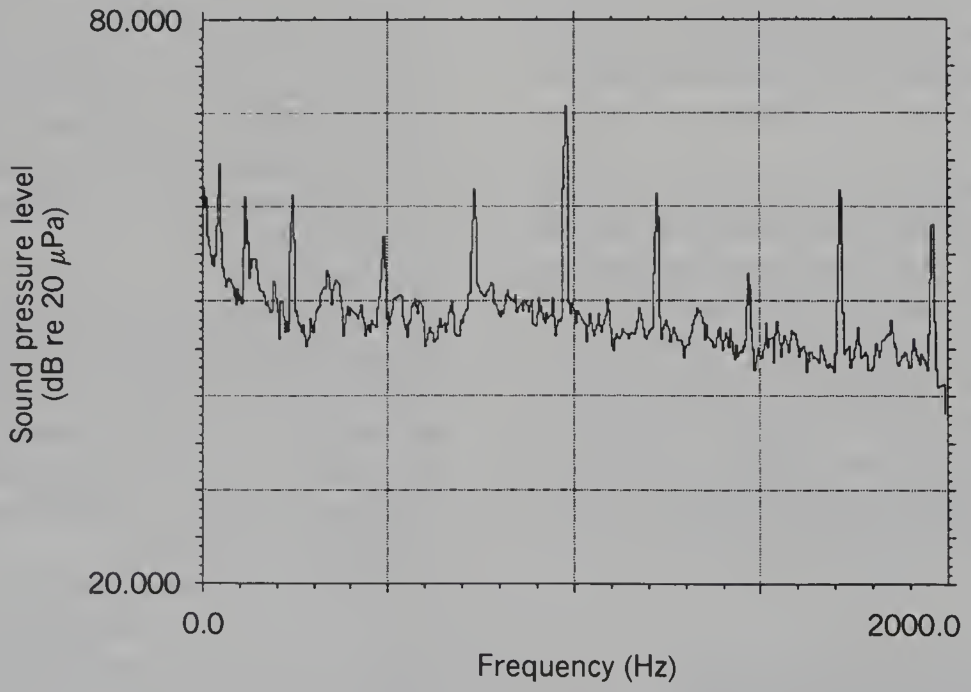


Fig. 2 Spectrum of electric fan noise.

computers. The signal from such a noise source is shown in Fig. 1. The frequency spectrum of the electric fan noise is shown in Fig. 2. Note that the spectrum consists of several distinct peaks (the deterministic part of the signal) superimposed on a broadband background (random) component. The peaks are produced by the repetitive processes of the fan, including the shaft rotational frequency (first peak, 49 Hz), the twice-line-frequency electric motor hum (second peak, 120 Hz), the blade passage frequency–shaft rotational frequency times number of fan blades (third peak at 245 Hz), and harmonics of the blade passage frequency (peaks at 490, 735, 980 Hz, etc.). The broadband background portion of the spectrum is due to the noise induced by the random turbulent flow of air through the fan and fan housing.

Signal processing is an important tool in acoustics and noise control. However, like all tools, signal processing has limitations and must be used correctly. The previous example shows that one must have an understanding of the physical mechanisms in order to accurately interpret signal processing data. Second, signal processing techniques can only analyze and display physical effects, for example, those that can be measured with a transducer, while the subjective effects of sound must be established by a skilled professional. Figure 1*b*, for example, shows that most of the sound energy is contained in the broadband random portion of the spectrum, and this is what one would strive to reduce if the objective is to reduce the overall sound emission of the fan. However, it is the deterministic components of the signal, which contain little energy, that one must reduce to make the fan less objectionable from the standpoint of noise.

100

STATISTICAL THEORY OF ACOUSTIC SIGNALS

ALLAN G. PIERSOL

1 INTRODUCTION

This chapter details the basic statistical functions used to describe deterministic, transient, and random acoustic signals and summarizes the applications of these functions to general acoustical problems. The first section following this introduction addresses magnitude domain descriptions (average values and probability functions); the second section covers time-domain descriptions (correlation functions); and the third section details frequency-domain descriptions (spectral functions). The fourth section discusses joint-function descriptions (cross-correlation and cross-spectral functions), while the final section summarizes the statistical sampling errors associated with the estimation of the various functions from time history measurements of acoustic signals. All functions are defined in analog terms, but the equations can be readily converted to discrete terms by simply replacing the continuous signal, $x(t)$; $0 \le t \le T$, by the discrete-valued signal, $x(n\,\Delta t)$; $n = 1, 2, \ldots, N$, where Δt is the time interval between the discrete values and $T = N\,\Delta t$. The actual computations of the various functions are addressed in Chapter 101 and the cited references. Analysis hardware and software are covered in Part XVII.

2 MAGNITUDE-DOMAIN PARAMETERS AND FUNCTIONS

A *stationary signal* is either a deterministic or random signal whose average properties do not vary with translations in time, for example, all periodic signals are stationary. Magnitude-domain descriptions of acoustic signals are generally limited to stationary acoustic signals, although the various descriptions can sometimes be applied to a nonstationary acoustic signal by assuming the signal is piecewise stationary and computing the desired parameters and functions over each of a series of contiguous, short time segments.[1] The following definitions apply to a stationary acoustic signal $x(t)$ defined over the time interval $0 \le t \le T$.

2.1 Mean and rms Values

In general, various different average values (moments) of a stationary signal $x(t)$ might be computed, although the *mean value* μ_x and the *standard deviation* σ_x (or the *rms value* ψ_x) are usually the only average values of interest. These average values are defined by

$$\mu_x = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t)\,dt \tag{1}$$

$$\sigma_x = \lim_{T \to \infty} \left\{ \frac{1}{T} \int_0^T [x(t) - \mu_x]^2\,dt \right\}^{1/2} \tag{2}$$

$$\psi_x = \lim_{T \to \infty} \left\{ \frac{1}{T} \int_0^T x^2(t)\,dt \right\}^{1/2} \tag{3}$$

where T is the linear averaging time. It can be shown[1] that the three quantities defined in Eqs. (1)–(3) are interrelated by

$$\psi_x^2 = \mu_x^2 + \sigma_x^2. \tag{4}$$

Hence, a knowledge of any two quantities determines the

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

third. The mean value μ_x defines the central tendency of the signal, while the standard deviation σ_x defines the dispersion, each with the same units as the signal (usually pressure in pascals for acoustic signals). The rms value is a function of both central tendency and dispersion. For most acoustic measurements, the transducer does not sense the static or DC component of the signal. For this case, the mean value of the signal is zero ($\mu_x = 0$) and $\psi_x = \sigma_x$, as will be assumed henceforth.

The limiting operations on T in Eqs. (1)–(3) cannot be accomplished in practice. For the case of stationary deterministic (periodic) signals, there will be little or no error in the computed average values as long as either $T = iT_p$; $i = 1, 2, 3, \ldots$, or $T \gg T_p$, where T_p is the period of the signal. For stationary random signals, however, there will be a statistical sampling (random) error in the computed average values that diminishes with an increasing value of T, as well as with an increasing value of the signal bandwidth B^2. The statistical sampling errors for mean and rms value estimates are summarized in Section 7.

2.2 Probability Distribution Functions

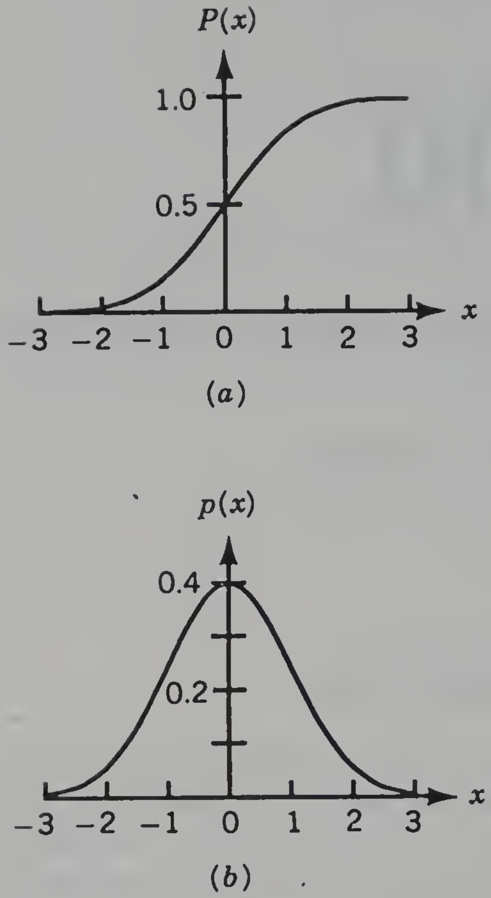
For periodic signals, which can be described by an equation, the exact value of the signal at any future time is determined simply by substituting the time of interest into the equation. For stationary random signals, however, no equation exists, so the value of the signal at a future time can be predicted only in probability terms. The basic function used to describe the potential value of a random signal at a future time is called the *probability distribution function* $P(x)$, which defines the probability that the signal $x(t)$ will be equal to or less than a specific value x at any arbitrary time t, that is,

$$P(x) = \text{Prob}[x(t) \leq x]. \tag{5}$$

From a computational viewpoint, the probability distribution function is given by

$$\lim_{T \to \infty} \frac{T_{x(t) \leq x}}{T}, \tag{6}$$

where $T_{x(t) \leq x}$ is the time spent by the signal during the duration T with a magnitude of $x(t) \leq x$. The plot of a typical probability distribution function for a stationary random signal with a mean value of zero and a standard deviation of unity is illustrated in Fig. 1*a*. The units of the plot are probability versus instantaneous magnitude (usually pressure in pascals for acoustic signals). The interpretation of probability distribution functions is straightforward, namely, $P(x)$ defines the probability that the magnitude of the signal at a future time (assuming the signal remains stationary) will have a value equal to or less than x.



(*a*)

(*b*)

Fig. 1 Typical probability distribution and density functions for random signal ($\mu_x = 0, \sigma_x = 1$): (*a*) distribution function; (*b*) density function.

2.3 Probability Density Functions

Referring to Fig. 1*a*, the value of a probability distribution function always starts at zero for magnitudes approaching $-\infty$ and ends at unity for magnitudes approaching $+\infty$; it is the manner in which the function varies from zero to unity that differentiates random signals. Hence, it is common to look at the slope of the probability distribution function, which gives the *probability density function* $p(x)$, defined as[2]

$$p(x) = \frac{dP(x)}{dx}. \tag{7}$$

From a computational viewpoint, the probability density function is given by

$$p(x) = \lim_{T \to \infty} \lim_{\Delta x \to 0} \frac{T_{x(t) \in \Delta x}}{T\,\Delta x}, \tag{8}$$

where $T_{x(t) \in \Delta x}$ is the time spent by the signal during the duration T in a narrow magnitude window Δx centered at the magnitude value x, and it is assumed that T approaches infinity faster than Δx approaches zero. The plot of the probability density function associated with the probability distribution function in Fig. 1*a* is shown in Fig. 1*b*. The units of the plot are probability density (probability per unit magnitude) versus instantaneous magnitude (usually pressure in pascals for acoustic signals). From the viewpoint of interpretations, the integral of the probability density function (the area under the function) between any two magnitude values, x_1 and x_2, defines the probability that the magnitude of the signal at a future time (assuming the signal remains stationary) will have a value between x_1 and x_2; that is,

$$\int_{x_1}^{x_2} p(x)\,dx = \text{Prob}[x_1 \le x(t) \le x_2]. \qquad (9)$$

It follows from Eq. (5) that

$$\int_{-\infty}^{x} p(x')\,dx' = P(x) \quad \text{and} \quad \int_{-\infty}^{\infty} p(x)\,dx = 1. \qquad (10)$$

Since the limiting operations in Eq. (8) cannot be accomplished in practice, the computation of probability density functions will involve both a statistical sampling error due to the finite measurement duration T, and a magnitude resolution bias error due to the finite magnitude interval Δx^2. These statistical sampling and bias errors are summarized in Section 7.

2.4 Normal (Gaussian) Probability Density Function

Due to the practical implications of the Central Limit Theorem in statistics,[3] there is a strong tendency for a wide class of stationary random signals, including random acoustic signals, to have a probability density function that is closely approximated by the *normal*, or *Gaussian, probability density function*, given by[4]

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[\frac{-(x-\mu_x)^2}{2\sigma_x^2}\right], \qquad (11)$$

where μ_x and σ_x are the mean value and standard deviation, respectively, of the signal, as defined in Eqs. (1) and (2). The probability distribution and density functions shown in Fig. 1 are of the normal form. There sometimes are significant deviations from normality in random acoustic signals, particularly when the random signals are mixed with periodic components[2] or when nonlinearities are involved in the mechanisms producing the signals,[5] which may have a direct bearing on their interpretations and applications.

3 TIME-DOMAIN FUNCTIONS

For a deterministic signal, the linear relationship between the signal values at two different times, t and $t+\tau$, can be determined directly from the equation for the signal. For stationary random signals, however, this linear relationship is defined by the *autocorrelation function* $R_{xx}(\tau)$, which is given by[2]

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T-\tau} \int_0^{T-\tau} x(t)x(t+\tau)\,dt. \qquad (12)$$

The plot of a typical autocorrelation function for a wide bandwidth random signal is shown in Fig. 2. The units of the plot are squared magnitude (usually squared pressure for acoustic signals, i.e., squared pascals) versus time displacement in seconds. For stationary random signals, the inability to accomplish the limiting operation in Eq. (12) will produce a statistical sampling error in the computed autocorrelation values,[2] as summarized in Section 7.

It is seen from Eq. (12) that autocorrelation functions are always even, that is, $R_{xx}(-\tau) = R_{xx}(\tau)$. Also, it is clear from Eq. (3) that $R_{xx}(0) = \psi_x^2$. For the usual case of acoustic signals where $\mu_x = 0$, $R_{xx}(0) = \sigma_x^2$. For sta-

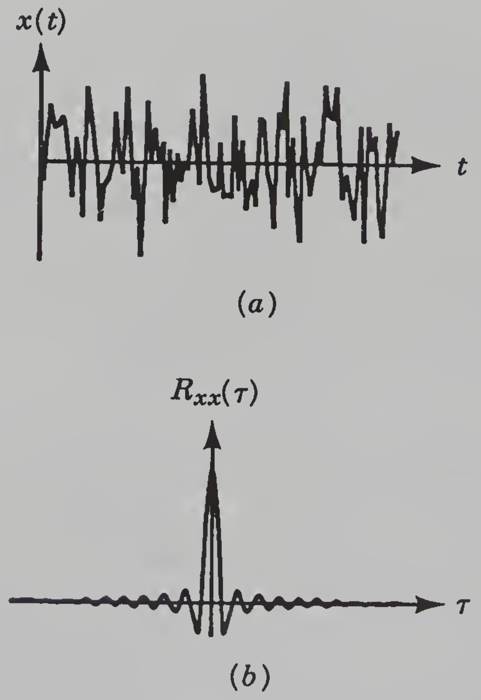


Fig. 2 Typical autocorrelation function for wide-bandwidth random signal: (*a*) signal; (*b*) autocorrelation function.

tionary random signals, the magnitude of the autocorrelation function will be less than $R_{xx}(0)$ as τ increases in either the positive or negative direction. The envelope of the autocorrelation function can be directly interpreted as a measure of how well future values of the signal can be predicted based on a knowledge of past values.[2,6] However, in most cases, the information contained in the autocorrelation function is more easily interpreted in terms of the Fourier transform of the autocorrelation function, called the power spectral density function discussed in Section 4.2.

4 FREQUENCY-DOMAIN FUNCTIONS

The most common and useful presentation for individual acoustic signals is some form of frequency decomposition. As noted in other chapters of this handbook, a spectral description in terms of rms values (usually expressed in decibels referenced to 20 μPa in air and 1 μPa in water) in one-third-octave frequency bands is a universal standard for the presentation of general acoustical data.[7] For detailed studies, however, a more comprehensive spectral description of acoustic signals is often required.

4.1 Spectral Description of Periodic Signals

Given a periodic signal, $x(t)$, $0 \le t \le T$, where $T = iT_P$, $i = 1, 2, 3, \ldots$, the frequency decomposition of the signal is given by the *linear (line) spectrum* $|P_x(f)|$, which is defined for nonnegative frequencies (a one-sided spectrum) by

$$|P_x(f)| = \begin{cases} \dfrac{2}{T}\,|X(f,T)|, & f > 0, \\[2mm] \dfrac{1}{T}\,|X(f,T)|, & f = 0, \\[2mm] 0, & f < 0, \end{cases} \tag{13}$$

where $|X(f,T)|$ is the magnitude of the finite Fourier transform of $x(t)$, $0 \le t \le T$, defined by

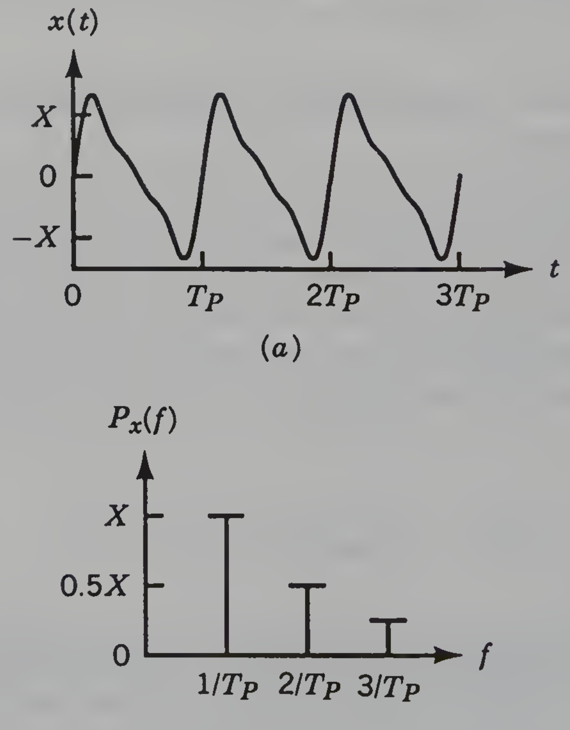
$$\begin{aligned} X(f,T) &= \int_0^T x(t) e^{-j2\pi ft}\, dt \\ &= \int_0^T x(t)\cos(2\pi ft)\, dt \\ &\quad - j\int_0^T x(t)\sin(2\pi f t)\, dt, \\ &\qquad -\infty \le f \le \infty. \end{aligned} \tag{14a}$$

For a discrete-valued signal, $x(n\,\Delta t)$, $n = 0, 1, \ldots, N-1$ (it is convenient to index the signal values for Fourier transforms from $n = 0$ rather than $n = 1$),

$$X(k\,\Delta f) = \Delta t \sum_{n=0}^{N-1} x(n\,\Delta t)\ \exp[-j2\pi kn/N], \qquad k = 0, 1, \ldots, N-1. \tag{14b}$$

where $\Delta f = 1/(N\ \Delta t)$. The discrete-valued formulation in Eq. (14b) is usually referred to as a *discrete Fourier transform* (DFT). The DFT of a discrete-valued signal can be computed very rapidly using algorithms commonly referred to as *fast Fourier transform* (FFT) procedures,[8] as discussed in Chapter 101.

In Eq. (14), note that $X(f,T)$ or $X(n\,\Delta f)$ is defined for both positive and negative frequencies (in the discrete-valued formulation, the spectral components at $k > \frac{1}{2}N$ correspond to negative frequency components). This is why a factor of 2 appears in the linear spectrum defined for positive frequencies only in Eq. (13). A typical plot of the linear spectrum for a periodic signal is presented in Fig. 3. The units of the plot are magnitude (usually pressure in pascals for acoustic signals) versus frequency in hertz. Assuming $T = iT_P$, $i = 1, 2, 3, \ldots$, the spectrum $|P_x(f)|$ in Eq. (13) represents the magnitude of the Fourier series coefficients of $x(t)$. In most cases, the phase of the Fourier coefficients is omitted, although phase information can be retained from Eq. (14)



(*a*)

(*b*)

Fig. 3 Typical Fourier (line) spectrum for a periodic signal: (*a*) signal; (*b*) spectrum.

if desired (the phase for the Fourier coefficients plotted in Fig. 3 is zero).

4.2 Spectral Description of Stationary Random Signals

Given a stationary random acoustic signal $x(t)$, $0 \le t \le T$, the frequency decomposition of the signal is given by the *power spectral density function* $G_{xx}(f)$ (also called the *power spectrum*, *autospectral density function*, or *autospectrum*), which is defined for nonnegative frequencies (a one-sided spectrum) by

$$G_{xx}(f) = \begin{cases} \lim\limits_{T \to \infty} \dfrac{2}{T} E|X(f,T)|^2, & f > 0, \\ \lim\limits_{T \to \infty} \dfrac{1}{T} E|X(f,T)|^2, & f = 0, \\ 0, & f < 0, \end{cases} \tag{15}$$

where $X(f,T)$ is the finite Fourier transform defined in Eq. (14), and $E[\cdot]$ denotes the expected value, that is, an average over an infinite ensemble of squared Fourier transform computations from statistically independent blocks of data.

It can be shown[2] that the power spectral density function is equal to the Fourier transform of the autocorrelation function defined in Eq. (12); that is,

$$\begin{aligned} G_{xx}(f) &= 2\int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j2\pi f\tau}\, d\tau \\ &= 2\int_{-\infty}^{\infty} R_{xx}(\tau)\cos(2\pi f\tau)\, d\tau, \\ &\qquad f \ge 0, \end{aligned} \tag{16}$$

where the factor of 2 appears because $G_{xx}(f)$ is a one-sided spectrum, and the second equality occurs because $R_{xx}(\tau)$ is an even function of τ. A typical plot of the power spectrum for a stationary random signal with a wide bandwidth is shown in Fig. 4. The units of the plot are squared magnitude per unit frequency (usually squared pressure per hertz for acoustic signals, e.g., squared pascals per hertz) versus frequency in hertz. It should be mentioned that some academic references define the power spectrum in terms of angular frequency, where the units are squared magnitude per radian per second versus radian per second.

The interpretation of the power spectrum is facilitated by two important properties. First, the integral of the power spectrum (the area under the function) between any two frequencies, f_1 and f_2, defines the mean-square value (the square of the rms value) of the signal between f_1 and f_2; that is,

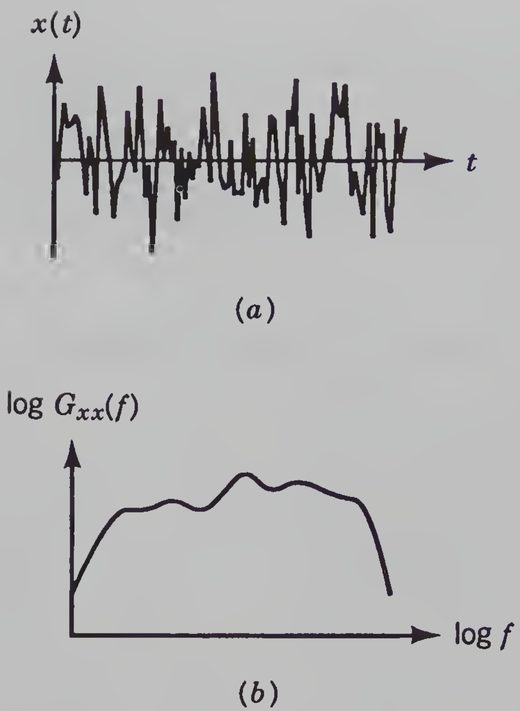


Fig. 4 Typical power spectral density function for wide-bandwidth random signal: (*a*) signal; (*b*) spectrum.

$$\int_{f_1}^{f_2} G_{xx}(f)\, df = \psi_x^2[f_1 \le f \le f_2], \tag{17}$$

where $\psi_x^2 = \sigma_x^2$ if either $\mu_x = 0$ or $f_1 > 0$. It follows that

$$\int_0^{\infty} G_{xx}(f)\, df = \psi_x^2 = R_{xx}(0). \tag{18}$$

Second, given two or more statistically independent signals with power spectra $G_{ii}(f)$, $i = 1, 2, \ldots, M$, the power spectrum for the sum of the signals is the sum of their individual power spectra; that is,

$$G_{xx}(f) = \sum_{i=0}^{M} G_{ii}(f). \tag{19}$$

Referring back to Eq. (15), neither the expected value operation nor the limiting operation on T can be accomplished in practice. However, the expected value can be approximated by averaging the individual power spectra computed from each of a finite number of disjoint (statistically independent) blocks of data, often created by subdividing one long measurement into n_d contiguous segments, each of finite length T (see Chapter 101).

The number of segments n_d for the averaging operation determines the statistical sampling error of the resulting power spectrum estimate,[2] while the inability to accomplish the limiting operation on the duration T produces a frequency resolution bias error[2,9] in the estimate. These statistical sampling and bias errors are summarized in Section 7.

4.3 Spectral Description of Nonstationary Signals

A *nonstationary signal* is a signal that is ongoing but whose average properties vary with translations in time. A spectral representation for a nonstationary acoustic signal is sometimes generated by simply computing the Fourier transform in Eq. (14) over short, contiguous time segments, each of duration T, and plotting the results with time as the ordinate axis, frequency as the abscissa axis, and magnitude as the darkness of the plotted data point. Such plots are sometimes called *spectrograms.* A time-varying spectral representation for a nonstationary signal can also be produced by repeatedly computing a short time-averaged rms value for the output of each of a set of contiguous narrow bandpass filters.[1,6] However, a more rigorous description of the spectral content of nonstationary acoustic signals is given by the *instantaneous power spectral density function* $G_{xx}(f,t)$ (also called the *instantaneous power spectrum*), which is defined for nonnegative frequencies (a one-sided spectrum) by[2]

$$G_{xx}(f,t)=\begin{cases} 2\displaystyle\int_{-\infty}^{\infty} R_{xx}(\tau,t)e^{-j2\pi f\tau}\,d\tau, & f>0,\\ \displaystyle\int_{-\infty}^{\infty} R_{xx}(\tau,t)e^{-j2\pi f\tau}\,d\tau, & f=0\\ 0, & f<0,\end{cases} \tag{20}$$

where $R_{xx}(\tau,t)$ is the *instantaneous autocorrelation function* given by

$$R_{xx}(\tau,t)=E\left[x\left(t-\frac{\tau}{2}\right)x\left(t+\frac{\tau}{2}\right)\right], \tag{21}$$

where, again, $E[\cdot]$ denotes the expected value, that is, an average over an infinite ensemble of product computations from statistically independent blocks of data. For deterministic data, there are no statistical sampling errors in the product computation and, hence, the expected value operation is not needed. The instantaneous power spectrum in this case is often called the *Wigner–Ville distribution*. For random signals, some form of averaging over small time and/or frequency increments is often employed to suppress statistical sampling errors in instantaneous power spectra estimates.[1,6]

The interpretation of the instantaneous power spectrum at any time t is the same as discussed for the time-averaged power spectrum in Section 4.2, except the instantaneous power spectrum can have negative values at certain combinations of time and frequency. In all cases, however, the integral of the instantaneous power spectrum over frequency at any time yields the mean-square value of the signal at that time, that is,

$$\int_0^{\infty} G_{xx}(f,t)\,df=\psi_x^2(t). \tag{22}$$

Furthermore, the integral of the instantaneous power spectrum over time yields

$$\int_0^{\infty} G_{xx}(f,t)\,dt=E_{xx}(f), \tag{23}$$

where $E_{xx}(f)$ is the energy spectral density function defined in Section 4.4.

4.4 Spectral Description of Transient Signals

There are two related functions used to describe transient acoustic signals in the frequency domain; (a) the Fourier spectrum, which is appropriate for deterministic transients and (b) the energy spectral density function, which is appropriate for random transients.

Fourier Spectrum For the case of a deterministic transient signal $x(t)$, $0 \le t \le T$, a spectral description of the transient is provided by its *Fourier spectrum* $F_x(f)$, which is defined for nonnegative frequencies (a one-sided spectrum) by

$$F_x(f)=\begin{cases} 2X(f,T), & f>0,\\ X(f,T), & f=0,\\ 0, & f<0,\end{cases} \tag{24}$$

where $X(f,T)$ is the finite Fourier transform defined in Eq. (14). Assuming the block duration T equals or exceeds the duration of all significant values of the transient, the finite Fourier transform will essentially yield sample values of the exact Fourier transform of the transient $x(t)$. This follows because the values of the transient outside the time interval of the computation must be zero and, hence, all values of the transient are known from

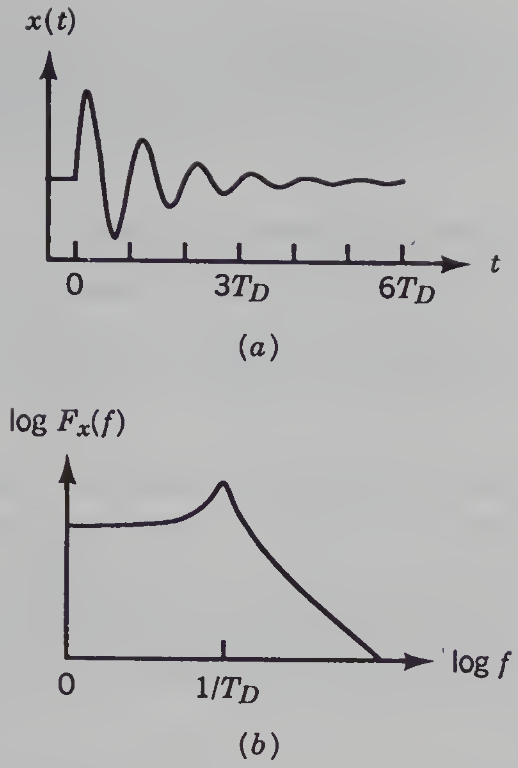


Fig. 5 Typical Fourier spectrum for transient signal: (*a*) signal; (*b*) spectrum.

minus to plus infinity. For many applications, the phase portion of the Fourier transform is ignored and only the Fourier magnitude spectrum, $|F_x(f)|$, is presented.

A typical plot of the Fourier spectrum for a deterministic transient is shown in Fig. 5. The units of the plot are the product of magnitude and time (usually pascals by seconds for acoustic signals) versus frequency in hertz. The interpretation of the Fourier spectrum for a transient signal is similar to that of a linear spectrum for the periodic signal discussed in Section 4.1, except the Fourier spectrum is theoretically continuous over all frequencies, rather than a line spectrum at discrete frequencies.

Energy Spectral Density Function For the case of a transient signal $x(t)$, $0 \le t \le T$, that is random in character, it is common to describe the frequency content of the signal in terms of an *energy spectral density function* $E_{xx}(f)$ (also called the *energy spectrum*), which is defined for nonnegative frequencies (a one-sided spectrum) by

$$E_{xx}(f) = \begin{cases} 2E[|X(f,T)|^2], & f > 0, \\ E[|X(f,T)|^2], & f = 0, \\ 0, & f < 0, \end{cases} \qquad (25)$$

where $X(f,T)$ is the finite Fourier transform defined in Eq. (14), and $E[\cdot]$ denotes the expected-value operation, as discussed for the power spectrum in Eq. (15). In this case, however, the expected value is approximated by averaging over the energy spectra computed from repeated transient measurements. As for the Fourier spectrum in Eq. (24), if the duration T in Eq. (25) equals or exceeds the duration of all significant values of the transient, the finite Fourier transform will essentially yield sample values of the exact Fourier transform of the transient $x(t)$. Plots of energy spectra often look much like the plots of power spectra, except the units of the energy spectrum plot are squared magnitude per unit frequency multiplied by time (usually squared pressure per hertz multiplied by time for acoustic signals, e.g., pascals squared by seconds per hertz) versus frequency in hertz. The interpretations of the energy spectrum and the statistical sampling error in its estimation are similar to those for the power spectrum, except the integral of an energy spectrum over frequency yields a mean-square value multiplied by time (e.g., pascals squared by seconds), rather than just a mean-square value.

5 JOINT FUNCTIONS

Many applications of acoustic data are served by the functions for individual signals discussed in Sections 2–4. For more advanced applications involving random acoustic signals, however, joint functions of two or more signals may be required. There are some applications where it is convenient to present the desired information in the time domain using cross-correlation and related functions. However, it is generally more common to work directly in the frequency domain using cross-spectral density and related functions.

5.1 Time-Domain Functions

There are three related time-domain descriptions of stationary random signals that are of common interest: (a) cross-correlation functions, (b) correlation coefficient functions, and (c) unit impulse response functions.

Cross-Correlation Functions Given two stationary random signals $x(t)$ and $y(t)$, $0 \le t \le T$, any linear relationship between these two signals will be extracted by the *cross-correlation function* $R_{xy}(\tau)$, defined by

$$R_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{T - \tau} \int_0^{T-\tau} x(t)y(t+\tau)\,dt. \qquad (26)$$

A typical plot of a cross-correlation function between two correlated random signals with wide bandwidths is shown in Fig. 6. In this plot, the signals could be inter-

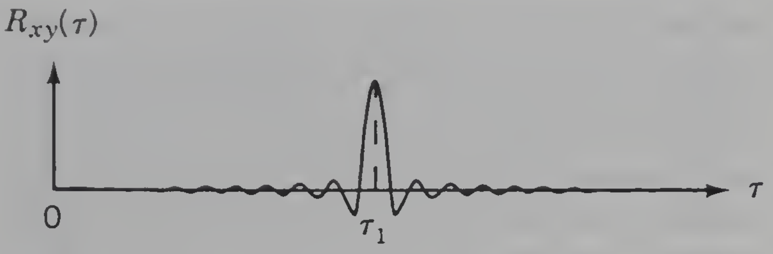


Fig. 6 Typical cross-correlation function between wide-bandwidth random signals representing measurements at two points along a propagating acoustic wave (propagation time $\tau = \tau_1$).

preted as two measurements along the path of a propagating acoustic wave, where the propagation time is τ_1 seconds. The units of the plot are the product of the magnitudes of $x(t)$ and $y(t)$ (usually pascals squared for acoustic signals) versus time displacement in seconds.

It can be shown[2] that the cross-correlation function is the inverse Fourier transform of the cross-spectral density function defined later in Eq. (28) and, hence, contains no information that is not available from a cross-spectrum. However, the time-domain character of the cross-correlation function may sometimes present information in a more desirable format for engineering evaluations. This is particularly true of applications involving time delay estimates,[6] as illustrated in Fig. 6. For stationary random signals, the inability to accomplish the limiting operation in Eq. (26) will produce a statistical sampling error in the computed cross-correlation values,[2] as summarized in Section 7.

Correlation Coefficient Functions For some applications, it is more convenient to estimate a normalized cross-correlation function, called the *correlation coefficient function* $\rho_{xy}(\tau)$, given by

$$\rho_{xy}(\tau) = \frac{R_{xy}(\tau)}{\sigma_x \sigma_y}, \tag{27}$$

where $x(t)$ and/or $y(t)$ has a zero mean value, and σ_x and σ_y are the standard deviations of $x(t)$ and $y(t)$, respectively, as defined in Eq. (2). The quantity $\rho^2_{xy}(\tau)$ is bounded by zero and unity and defines the fraction of the variance (the square of the standard deviation) of $y(t)$ that is linearly related to $x(t)$; that is, $\rho^2_{xy}(\tau) = 0$ means there is no linear relationship and $\rho^2_{xy}(\tau) = 1$ means there is a perfect linear relationship between $x(t)$ and $y(t)$ at the time displacement τ.

Unit Impulse Response Functions The *unit impulse response function* $h_{xy}(\tau)$ of a physical system is defined as the response of that system to a delta function input, that is, letting $y(t)$ be the response of a system to an input $x(t)$, $h_{xy}(\tau) = y(t)$ when $x(t) = \delta(t)$. The unit impulse response function is the inverse Fourier transform of the frequency response function defined in Section 5.2, and is more easily interpreted in that context.

5.2 Frequency-Domain Functions

There are three related frequency-domain descriptions of stationary random signals that are of common interest: (a) cross-spectral density functions, (b) coherence functions, and (c) frequency response functions.

Cross-Spectral Density Functions Given two signals $x(t)$ and $y(t)$, $0 \le t \le T$, any linear relationship between these two signals at various different frequencies will be extracted by the *cross-spectral density function* $G_{xy}(f)$ (also called the *cross-spectrum*), which is defined for nonnegative frequencies (a one-sided spectrum) by

$$G_{xy}(f) = \begin{cases} \lim\limits_{T\to\infty} \dfrac{2}{T}\, E[X^*(f,T)Y(f,T)], & f > 0, \\ \lim\limits_{T\to\infty} \dfrac{1}{T}\, E[X^*(f,T)Y(f,T)], & f > 0, \\ 0, & f < 0, \end{cases} \tag{28}$$

where $X^*(f,T)$ is the complex conjugate of the finite Fourier transform of $x(t)$, and $Y(f,T)$ is the finite Fourier transform of $y(t)$, as defined in Eq. (14). The expected-value notation $E[\cdot]$ has the same significance as discussed in Section 4.2. It should be mentioned that the cross-spectral density function is defined in some references as the average of $X(f,T)Y^*(f,T)$, which gives the complex conjugate of the definition in Eq. (28). Also, it can be shown[2] that the cross-spectral density function is equal to the Fourier transform of the cross-correlation function defined in Eq. (26); that is,

$$G_{xy}(f) = 2\int_{-\infty}^{\infty} R_{xy}(\tau) e^{-j2\pi f\tau}\, d\tau, \qquad f \ge 0, \tag{29a}$$

where the factor of 2 appears because $G_{xy}(f)$ is a one-sided spectrum.

The cross-spectral density function is generally a complex number that can be written as

$$G_{xy}(f) = C_{xy}(f) - jQ_{xy}(f), \tag{29b}$$

where the real part, denoted by $C_{xy}(f)$, is called the *coincident spectral density function* or *cospectrum*, and the imaginary part, denoted by $Q_{xy}(f)$, is called the *quadrature spectral density function* or *quad-spectrum.* The cross-spectrum may also be expressed in terms of magnitude and phase by

$$G_{xy}(f) = |G_{xy}(f)| e^{-j\theta_{xy}(f)}, \tag{29c}$$

where

$$|G_{xy}(f)| = [C_{xy}^2(f) + Q_{xy}^2(f)]^{1/2},$$

$$\theta_{xy}(f) = \tan^{-1}\left[\frac{Q_{xy}(f)}{C_{xy}(f)}\right]. \tag{29d}$$

A typical plot of the cross-spectral density function between two correlated random signals with wide bandwidths, having the same origin as discussed in Fig. 6, is presented in Fig. 7. The units of the plot are the product of the magnitudes of $x(t)$ and $y(t)$ per unit frequency (usually pascals squared per hertz for acoustic signals) versus frequency in hertz. The cross-spectrum may be interpreted as a measure of the linear dependence between two signals as a function of frequency, although the coherence function discussed next is more useful for this application. The statistical sampling and bias errors in cross-spectra estimates are summarized in Section 7.

Coherence Functions It is often convenient to normalize the cross-spectral density magnitude to obtain a quantity called the *coherence function* $\gamma_{xy}^2(f)$ (sometimes called coherency squared), given by

$$\gamma_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}, \tag{30}$$

where $G_{xy}(f)$ is the cross-spectrum defined in Eq. (28) and $G_{xx}(f)$ and $G_{yy}(f)$ are the power spectra defined in Eq. (15). The coherence function is bounded by $0 \leq \gamma_{xy}^2(f) \leq 1$ at all frequencies, and essentially identifies the fractional portion of linear dependence (or correlation) between two signals $x(t)$ and $y(t)$ as a function of frequency. Specifically, $\gamma_{xy}^2(f) = 0$ means there is no linear relationship, and $\gamma_{xy}^2(f) = 1$ means there is a perfect linear relationship between $x(t)$ and $y(t)$ at frequency f. For values between zero and unity, the coherence can be interpreted as the fractional portion of $G_{yy}(f)$ that can be determined from a knowledge of $G_{xx}(f)$. A plot of $\gamma_{xy}^2(f)$ is a dimensionless number versus frequency. The statistical sampling and bias errors in coherence function estimates are summarized in Section 7.

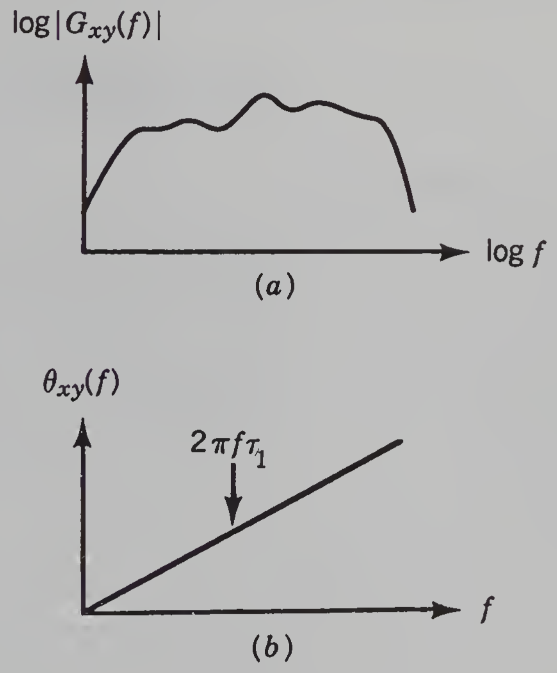


Fig. 7 Typical cross-spectral density function between wide-bandwidth random signals representing measurements at two points along a propagating acoustic wave (propagation time $\tau = \tau_1$): (*a*) magnitude; (*b*) phase.

Frequency Response Functions For two signals $x(t)$ and $y(t)$ representing the input and response, respectively, of a physical system, the ultimate goal of a cross-spectral density function computation is often the determination of a *frequency response function* $H_{xy}(f)$ for the system, which is given by[2]

$$H_{xy}(f) = \frac{G_{xy}(f)}{G_{xx}(f)}, \tag{31a}$$

where $G_{xy}(f)$ is the cross-spectrum defined in Eq. (28), and $G_{xx}(f)$ is the power spectrum defined in Eq. (15). Noting that $G_{xy}(f) = C_{xy}(f) - jQ_{xy}(f)$, a complex quantity, the frequency response function can be written in terms of a magnitude and phase, as follows:

$$H_{xy}(f) = |H_{xy}(f)| e^{-j\phi_{xy}(f)}, \tag{31b}$$

where

$$|H_{xy}(f)| = \frac{[C_{xy}^2(f) + Q_{xy}^2(f)]^{1/2}}{G_{xx}(f)},$$

$$\phi_{xy}(f) = \tan^{-1}\left[\frac{Q_{xy}(f)}{C_{xy}(f)}\right]. \tag{31c}$$

The frequency response function of a physical system is the Fourier transform of the unit impulse response of the system defined in Section 5.1; that is,

$$H_{xy}(f) = \int_0^\infty h_{xy}(\tau) e^{-j2\pi f\tau}\, d\tau. \tag{31d}$$

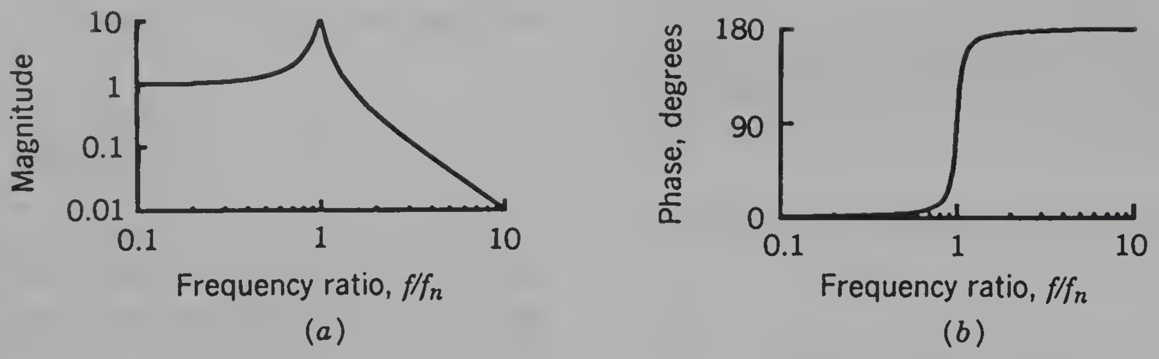


Fig. 8 Typical frequency response function for single-degree-of-freedom system ($\zeta = 0.05$): (*a*) magnitude; (*b*) phase.

A typical plot of the frequency response function for a single degree-of-freedom system with 5% damping (see Chapter 63) is shown in Fig. 8. The units of the plot are the ratio of the magnitudes of $y(t)$ over $x(t)$ (e.g., metres per newton) versus frequency in hertz. Frequency response functions are often called transfer functions, even though the transfer function is a more detailed description of the physical system given by the Laplace

TABLE 1 Summary of Estimation Errors for Common Signal Analysis Estimates

Parameter or Function, ϕ	Estimation Formula, $\hat{\phi}$	Normalized Random Error, $\epsilon_r(\hat{\phi})$	Normalized Bias Error, $\epsilon_b(\hat{\phi})$
μ_x	$\frac{1}{T}\int_0^T x(t)\,dt$	$\frac{1}{\sqrt{2BT}}\left(\frac{\sigma_x}{\mu_x}\right)$	0
$\psi_x(\mu_x = 0)$	$\left\{\frac{1}{T}\int_0^T x^2(t)\,dt\right\}^{1/2}$	$\frac{1}{2\sqrt{BT}}$	0
$p(x)$	$\frac{T_{x(t)\in\Delta x}}{T\Delta x}$	$\frac{c}{\sqrt{2BT\Delta x}}$	$\frac{\Delta x^2}{24}\left(\frac{d^2p(x)/dx^2}{p(x)}\right)$
$R_{xx}(\tau)$	$\frac{1}{T-\tau}\int_0^{T-\tau} x(t)x(t+\tau)\,d\tau$	$\frac{1}{\sqrt{2BT}}\,[1+\rho_{xx}^{-2}(\tau)]^{1/2}$	0
$G_{xx}(f)$	$\frac{2}{n_dT}\sum_{i=1}^{n_d}\lvert X_i(f,T)\rvert^2$	$\frac{1}{\sqrt{n_d}}$	$\frac{\Delta f^2}{24}\left(\frac{d^2G_{xx}(f)/df^2}{G_{xx}(f)}\right)$
$R_{xy}(\tau)$	$\frac{1}{T-\tau}\int_0^{T-\tau} x(t)y(t+\tau)\,d\tau$	$\frac{1}{\sqrt{2BT}}\,[1+\rho_{xy}^{-2}(\tau)]^{1/2}$	0
$\lvert G_{xy}(f)\rvert$	$\left\lvert\frac{2}{n_dT}\sum_{i=1}^{n_d} X_i^*(f,T)Y_i(f,T)\right\rvert$	$\frac{1}{\lvert\gamma_{xy}(f)\rvert\sqrt{n_d}}$	$\frac{\Delta f^2}{24}\left(\frac{d^2\lvert G_{xy}(f)\rvert/df^2}{G_{xy}(f)}\right)$
$\gamma_{xy}^2(f)$	$\frac{\lvert\hat{G}_{xy}(f)\rvert^2}{\hat{G}_{xx}(f)\hat{G}_{yy}(f)}$	$\frac{\sqrt{2}(1-\gamma_{xy}^2(f))}{\lvert\gamma_{xy}(f)\rvert\sqrt{n_d}}$	$\frac{(1-\gamma_{xy}^2(f))^2}{\gamma_{xy}^2(f)n_d}$
$\lvert H_{xy}(f)\rvert$	$\frac{\lvert\hat{G}_{xy}(f)\rvert}{\hat{G}_{xx}(f)}$	$\frac{(1-\gamma_{xy}^2(f))^{1/2}}{\lvert\gamma_{xy}(f)\rvert\sqrt{2n_d}}$	—[a]

Symbols:

B = frequency bandwidth of signal assuming a uniform spectrum within B, Hz

T = total duration of signal or duration of data blocks for spectral analysis,

Δf = frequency resolution bandwidth for spectral analysis ($1/T$),

n_d = number of disjoint (statistically independent) data blocks for spectral analysis

Δx = magnitude resolution window width for probability density analysis, in units of x

c = coefficient with value $0.3 < c < 1$, depending on sampling rate

[a] See Refs 2 and 9.

transform of the unit impulse response function.[2] Also, the phase factor is often defined as the negative of the value given by Eq. (31c). Frequency response functions can be interpreted as the gain and phase of a filter that produces an output $y(t)$ due to an input $x(t)$. The statistical sampling error in a frequency response function estimate is summarized in Section 7.

6 MISCELLANEOUS FUNCTIONS

Other functions that may sometimes be required to properly describe acoustic signals for specialized applications include the following:

(a) Correlated or coherent output power functions for source identification studies[2,6]
(b) Higher order moments, such as the third (skewness) moment and the fourth (kurtosis) moment, for nonlinear signal studies[5]
(c) Higher order spectral density functions (e.g., bispectra) and coherence functions for nonlinear signal studies[5]
(d) Conditioned spectral density functions, including multiple and partial coherence functions, for correlated source studies[2,6]
(e) Hilbert transforms for the generation of envelope functions and causality studies[2]
(f) Wavelets for the analysis and simulation of transients[10]
(g) Parametric spectral analysis procedures for short duration measurements[11]

7 ESTIMATION ERRORS

The definitions for the various descriptive functions of random signals in Sections 2–5 involve limiting operations that cannot be accomplished in practice. It follows that the actual computation of each of these functions will provide only an estimate of the function, which will include a statistical sampling (random) error, and in some cases a bias error as well. It is convenient to quantify these estimation errors in normalized terms; that is,

$$\text{Random error} = \epsilon_r(\hat{\phi}) = \frac{\sigma_{\hat{\phi}}}{\phi},$$

$$\text{Bias error} = \epsilon_b = \frac{b_{\hat{\phi}}}{\phi}, \qquad (32)$$

where ϕ is a function of interest, $\hat{\phi}$ is an estimate of function ϕ, $\sigma_{\hat{\phi}}$ is the standard deviation of estimate $\hat{\phi}$, and $b_{\hat{\phi}}$ is the bias of the estimate $\hat{\phi}$.

The interpretation of the normalized errors is as follows. If $\epsilon_r(\hat{\phi}) = 0.1$, the standard deviation of the estimate $\hat{\phi}$ is 10% of the function ϕ being estimated. In terms of a confidence interval, it can be said that there is about 68% confidence that the estimate $\hat{\phi}$ is within 10% of the true value ϕ, assuming the estimate is unbiased. If $\epsilon_b(\hat{\phi})$ equals $+0.1$ or -0.1 (bias errors carry a sign), the estimate $\hat{\phi}$ is, on the average, 10% higher or 10% lower than the function ϕ being estimated (assuming $\phi \neq 0$).

Approximate expressions for the normalized estimation errors associated with the magnitudes of the functions defined in Sections 2–5 are summarized in Table 1. All formulas in Table 1 are presented in analog terms, but may be converted to discrete terms by substituting $x(n\,\Delta t)$ for $x(t)$ and $N\,\Delta t$ for T. The formulas are generally acceptable approximations for $\epsilon < 0.2$. For those cases where $\phi = 0$, the expressions for the normalized errors in Table 1 should be multiplied by ϕ to obtain error expressions in engineering units. See Ref. 2 for more general formulas and confidence intervals.

REFERENCES

1. H. Himelblau, *Handbook for Data Acquisition and Analysis*, IES-RP-DTE012.1, Institute of Environmental Sciences, Mount Prospect, IL, 1994.
2. J. S. Bendat and A. G. Piersol, *RANDOM DATA: Analysis and Measurement Procedures*, 2nd ed., Wiley, New York, 1986.
3. M. Loeve, *Probability Theory*, 4th ed., Springer-Verlag, New York, 1977.
4. J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*, Vol. 40: *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, 1982.
5. J. S. Bendat, *Nonlinear System Analysis from Random Data*, Wiley, New York, 1990.
6. J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, 2nd ed., Wiley, New York, 1993.
7. Anonymous, "American National Standard, Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters," ANSI Std S1.11-1986.
8. J. W. Cooly and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Math. Comput.*, Vol. 19, 1965, pp. 297–301.
9. H. Schmidt, "Resolution Bias Errors in Spectral Density, Frequency Response and Coherence Function Estimates," *J. Sound Vib.*, Vol. 101, 1985, pp. 347–427.
10. D. E. Newland, *Random Vibrations, Spectral and Wavelet Analysis*, 3rd ed., Wiley, New York, 1993.
11. S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

101

PRACTICAL CONSIDERATIONS IN SIGNAL PROCESSING

JOHN C. BURGESS

1 INTRODUCTION

Practical signal processing means planning to acquire valid data and selecting methods for data reduction to achieve useful results. Practical methods usually are based on an FFT (fast Fourier transform), an efficient implementation of the DFT (discrete Fourier transform). (As used in this chapter, DFT implies computation with an FFT.) The objective of this chapter is to provide insight into methods based on the DFT and their practical use.

This chapter includes discussion of the four principal factors that must be specified:

1. The maximum frequency $f_{\max}$ for information in a signal sets the sampling rate $f_s > 2f_{\max}$ (Section 2.5).
2. The minimum acceptable data record length T_r sets the minimum acceptable DFT block size $N = T_r f_s$. Also, T_r sets the nominal spectral resolution $\Delta f = 1/T_r = f_s/N$ (Sections 2.3 and 2.7).
3. A data window may be required to control leakage (Sections 2.6 and 4.2).
4. For random signals, block averaging normally is required to achieve desired statistical stability (Section 6.4). Lag windows may also be used (Section 6.6).

Additionally, where an excitation signal can be specified, chirps and choips can be designed with the twin advantages of low peak factor and energy concentration in a desired power band.[1]

There are three general methods for practical signal processing: covariance, parametric, and direct. The *covariance* method[2] has primary value for power spectrum estimation where very limited amounts of data are available (e.g., sun-spots). *Parametric* methods[3] have primary value where the amount of data is somewhat limited (e.g., quasi-stationary signals such as speech) and where models are desired.

Direct methods are based on the DFT. They can be used with limited or unlimited amounts of data and can reduce significantly the computation time required by covariance and parametric methods. However, a DFT *always* treats a single data record as if the underlying analog signal were periodic (deterministic). This means that care is required to recover correct values for periodic signal parameters (Section 4.3) and that the treatment of random signals must account for the deterministic nature of single DFTs (Section 6).

2 TRANSFORMS AND SPECTRA

Transform means any of several equations that can provide a spectrum given a signal. *Inverse transform* means the corresponding equations that can provide a signal given its spectrum. *Transform pair* means the two equations that define a transform and its inverse.

Different nomenclature has developed within three technologies—electrical engineering, mechanical engineering, and statistics—to describe equivalent signal processing artifacts (Table 1). *System*, *filter*, and *process* are synonyms. Commonly used filter synonyms are "transversal" for finite impulse response (FIR) and "recursive" for infinite impulse response (IIR).

Relationships among the four transform types used most often are summarized in Table 2. Fourier trans-

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

TABLE 1 Comparative Nomenclature for System Types[a]

Electrical Engineering	Mechanical Engineering	Statistics
Filter	System	Process
IIR filter (all pole)	System	AR process
FIR filter (all zero)	None	MA process
IIR filter (pole zero	System	ARMA process

[a] *Abbreviations:* IIR, infinite impulse response; FIR, finite impulse response; AR, autoregressive; MA, moving average; ARMA, autoregressive moving average.

Source: Adapted with permission from Ref. 4.

TABLE 2 Relationships among Transform Types

	Time Continuous: $x(t)$, Frequency Aperiodic	Time Discrete: $x(n)$, Frequency Periodic
Frequency continuous: $\mathbf{X}(f)$, time aperiodic	Laplace transform, Fourier transform	z-Transform
Frequency discrete: $\mathbf{X}(k)$, time periodic	Fourier series	DFT, FFT

Source: Adapted with permission from Ref. 4.

forms, Fourier series, and z-transforms are widely used for theoretical representations. Only finite Fourier transforms, Fourier series, and the DFT involve data records. Only the DFT is normally used to compute spectra.

2.1 Discrete Fourier Transform

The DFT equations normally used for signal processing,

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{X}(k) e^{j2\pi kn/N}, \qquad 0 \le n \le N-1, \quad \text{(1a)}$$

$$\mathbf{X}(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \qquad 0 \le k \le N-1, \quad \text{(1b)}$$

are based on a z-transform derivation,[5,6] where N is both the record length and the DFT size. Another transform pair often encountered is based on a Fourier series derivation.[7]

An FFT is an efficient computer algorithm used to compute a DFT.[5,6] There are many such algorithms,[8] referred to collectively as "the FFT." The computation time required for an N-point FFT is approximately $N \log_2 N$, while it is approximately N^2 for a direct DFT computation. The time savings are substantial with the relatively long data records typical of acoustical applications.[9] *DFT* represents Eq. (1b), and *IDFT* represents the inverse transform Eq. (1a).

2.2 Linear Spectrum

Equations (1) can be used to define a *discrete linear spectrum*,

$$\mathbf{X}(k) = X(k) e^{j\phi(k)}, \qquad (2)$$

where $X(k) = |\mathbf{X}(k)|$ represents a *magnitude spectrum*, $\phi(k)$ represents a *phase spectrum*, and k is an integer frequency index (Section 2.3). Spectra represented by Eq. (2) are called "linear," where necessary to distinguish them from "energy" or "power" spectra. Discrete linear and power spectra are also called "discrete Fourier" and "Fourier line" spectra.

Figures 1*b* and 2 present a data record. Figures 3 and 4 show two views of the linear spectrum of the data record that appears in Figs. 1*b* and 2 (signal "seen" by a DFT). Comparison of Figs. 4 and 5 shows that the signal analyzed usually is not the true signal. (Figure 5 shows the linear spectrum of the periodic signal with period T that appears in Fig. 1*a* ("true" signal).) A major analysis objective is to estimate true signal values from the DFT values (Section 4.3). The dimensions of spectra are shown in Table 3.

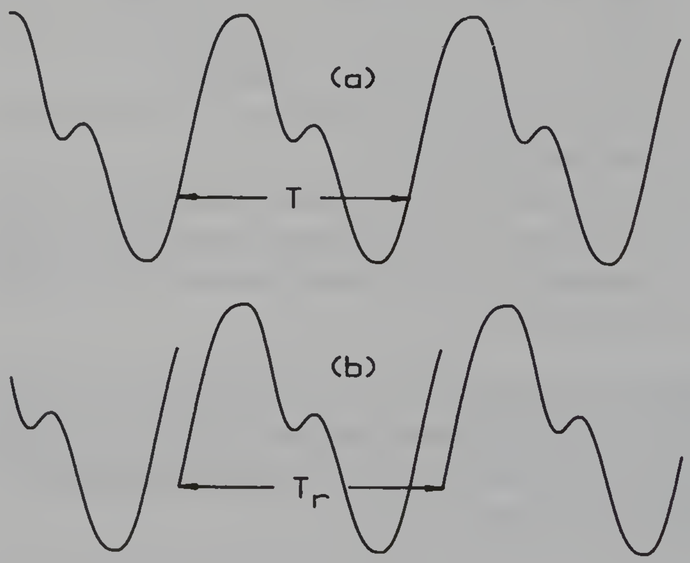


Fig. 1 (*a*) A periodic signal with period T. (*b*) A data record of length T_r and the periodic signal "seen" by repeating it.

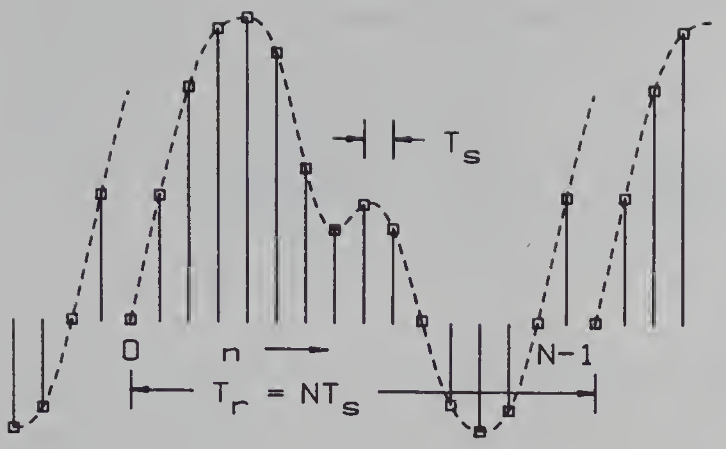


Fig. 2 Sampled periodic data record from Fig. 1*b*, $N = 16$.

All spectra discussed here are "two sided," meaning that they range over both negative and positive frequencies. This has particular value when analyzing the effects of data windows. The magnitude spectrum of a physical signal is an even function of frequency, and its phase spectrum is an odd function of frequency (Fig. 3). Since the negative-frequency part of such two-sided spectra can always be obtained from the positive part, it is common practice to show only the positive half, usually with normalized frequency [Eq. (9)] as the abscissa (Fig. 4). One-sided and two-sided magnitude spectra differ by a factor of 2; their phase spectra are the same.

For practical applications, power spectra are often given as "one sided," meaning that all the power is shown in the positive-frequency range.

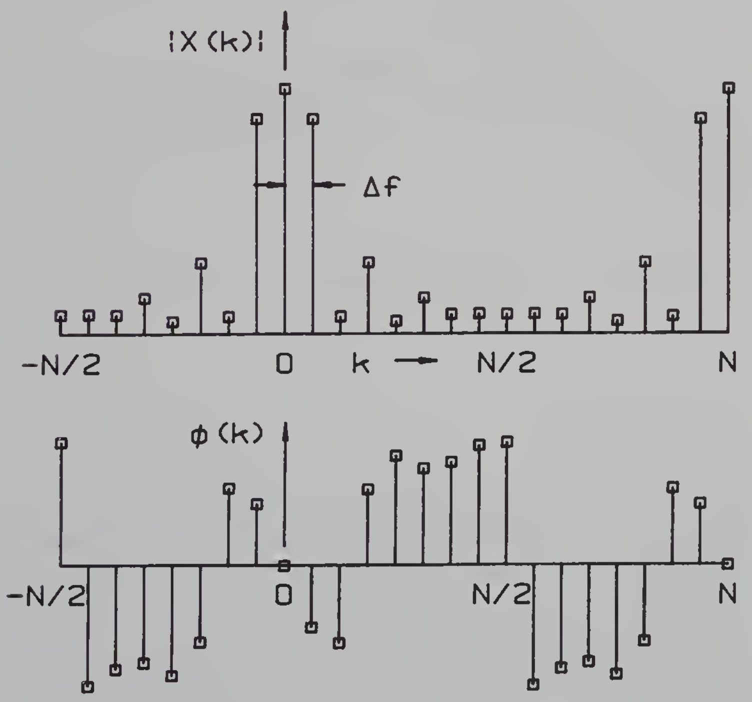


Fig. 3 Periodic discrete spectrum of the sampled data record in Fig. 2 calculated with a DFT.

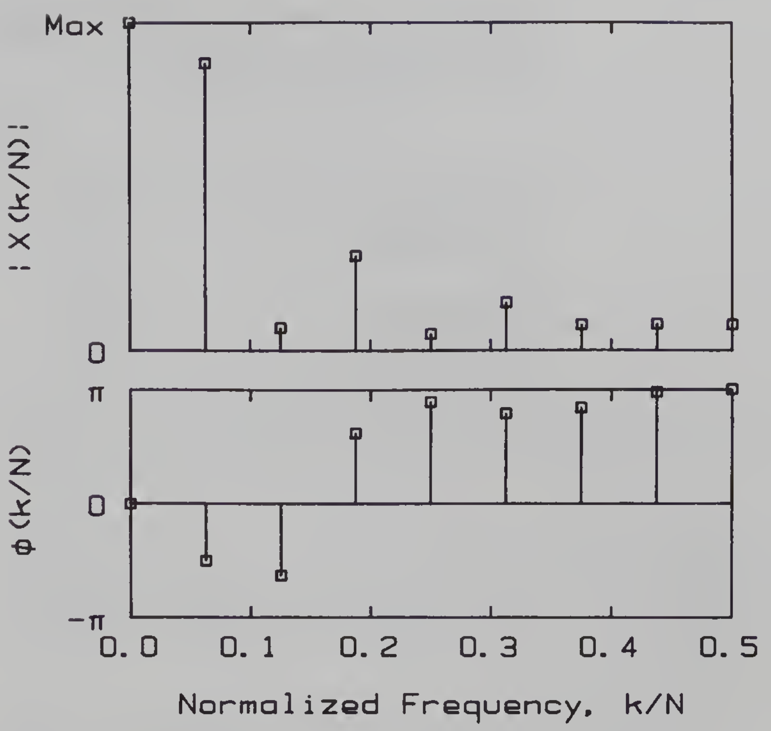


Fig. 4 Positive half of the two-sided spectrum from Fig. 3 with normalized frequency.

2.3 Signal and Spectrum Definitions

Discrete time and frequency are defined by

$$t(n) = n\,\Delta t = nT_s, \qquad (3a)$$

$$f(k) = k\,\Delta f, \qquad (3b)$$

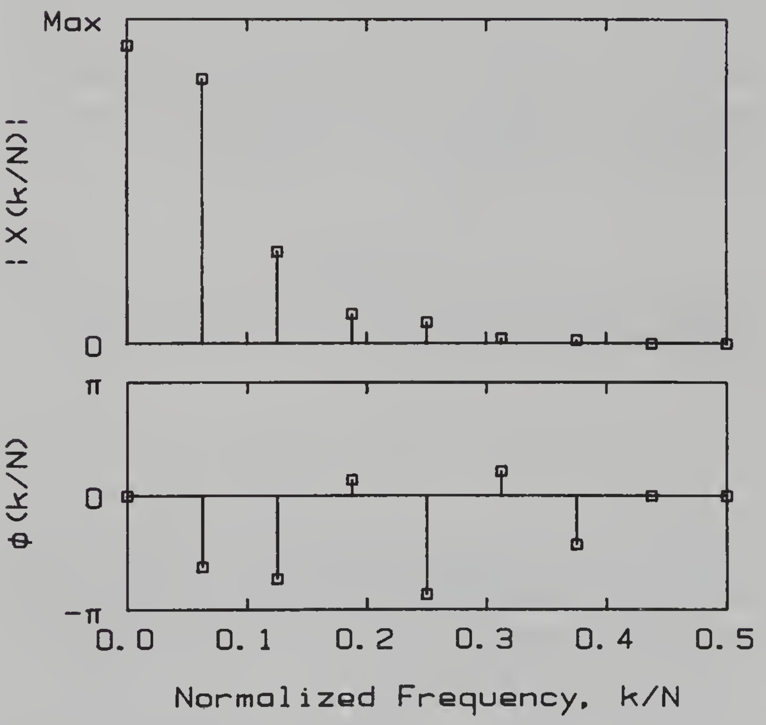


Fig. 5 Positive half of the two-sided spectrum of the periodic signal shown in Fig. 1*a* plotted with the same magnitude scale as Fig. 4.

TABLE 3 Dimensions of Linear, Energy, and Power Spectrum Magnitudes with Signal Dimensions of Volts

	Linear		Energy and Power	
Transform	Symbol	Dimensions	Symbol	Dimensions
Fourier series	$X(k)$	V	$X^2(k)$	V^2
DFT	$X(k)$	V	$X^2(k)$	V^2
z-Transform	$X(f)$	V	$X^2(f)$	V^2
Fourier transform	$X(f)$	V/Hz	$S(f)$	V^2/Hz

Source: Adapted with permission from Ref. 4.

where n and k are integers, called time and frequency *indices* (Figs. 2 and 3). The time between discrete data values, $T_s = \Delta t$, is the *sampling interval* (also called *sampling period* and *Nyquist interval*), and Δf is the spectral *frequency interval.*

The *sampling frequency* f_s (also called *sampling rate*) is the time rate at which discrete data values are taken, usually from an analog signal. The *Nyquist frequency* is defined here as

$$f_{Ny} = \tfrac{1}{2} f_s. \tag{4}$$

Nyquist rate was defined originally and is still used to mean *sampling rate*[10] (i.e., Nyquist *rate* means f_s and Nyquist *frequency* means $\frac{1}{2} f_s$. Only "Nyquist frequency" is used here. *Nyquist range* means the frequency range $-f_{Ny} \le f \le f_{Ny}$, or, alternately, the positive half range $0 \le f \le f_{Ny}$. Figures 1*b* and 2 show a signal as "seen" by a DFT. An analog data record $x(t)$ of duration T_r is taken from an original signal. The digital data record $x(n)$ of length N is related to T_r by Eq. (3a) and

$$T_r = NT_s, \tag{5}$$

where the sampling interval is

$$T_s = \frac{1}{f_s} = \frac{1}{2f_{Ny}}. \tag{6}$$

The analog data record extends from $t = 0$ to $t = T_r - \epsilon$, where ϵ is infinitesimally small, while unique discrete data values extend only from $t(0)$ to $t(N-1)$ (Fig. 2).

If T_r is measured in seconds,

$$\Delta f = f(1) = \frac{1}{T_r} = \frac{f_s}{N} \quad \text{Hz}. \tag{7}$$

The frequency corresponding to frequency index k is

$$f(k) = k\,\Delta f = \frac{k f_s}{N} \quad \text{Hz}. \tag{8}$$

The sampling frequency is often used to define a dimensionless *normalized frequency* (Figs. 4 and 5)

$$f_{\text{norm}} = \frac{f(k)}{f_s} = \frac{k}{N}. \tag{9}$$

The normalized frequency range $0 \le f_{\text{norm}} \le 0.5$ corresponds to the Nyquist range.

2.4 Periodicity and Circular Functions

Equations (1) make both $x(n)$ and $\mathbf{X}(k)$ periodic with period N (Figs. 2 and 3). Functions having this characteristic are called *circular.*

A computational advantage of this periodicity is that, after a data record has been acquired, sums taken over any sequential N values of $x(n)$ or $\mathbf{X}(k)$ are identical. For example,

$$\sum_{0}^{N-1} (\cdot) = \sum_{-N/2+1}^{N/2} (\cdot). \tag{10}$$

2.5 Aliasing

Shannon's information theorem[11] states that all information (signal components) in a continuous signal at frequencies $f \le f_{Ny}$ will be preserved when the signal is sampled at the constant rate $f_s \ge 2f_{Ny}$. *Aliasing* refers to the incorrect identification of components in a continuous signal at frequencies higher than f_{Ny} as low-frequency components in the sampled signal (Fig. 6). Components at frequencies higher than f_{Ny} are "aliased" to

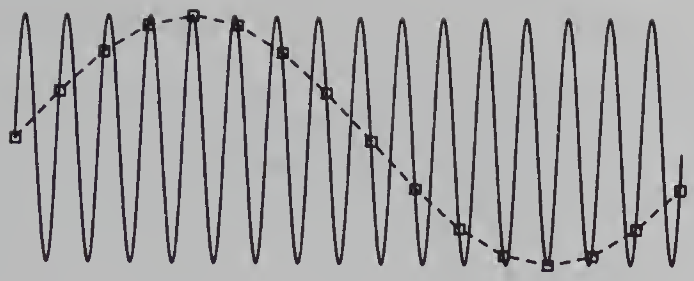


Fig. 6 An aliasing example.

lower frequencies as if the spectrum of the continuous signal were folded about the Nyquist frequency (f_{Ny} is sometimes called the "folding frequency"). For example, an 8-kHz tone sampled at a 10-kHz rate will appear as a 2-kHz signal:

> *There is only one way to prevent aliasing. Before sampling, be sure there are no components in the analog signal at frequencies higher than the Nyquist frequency.*

There are two practical ways to prevent aliasing:

- Use a sampling rate $f_s > 2f_{\max}$.
- Use an *anti-aliasing filter* (AAF) (Section 3.1).

2.6 Leakage and Data Windows

Leakage refers to spectral spreading in a transform that results from applying a transform to a finite-length data record. It does not result from sampling, and it does not depend on signal type.

Leakage is most easily observed and described with a sinusoidal signal. An infinitely long sampled signal with frequency f_m,

$$x(n) = \frac{A \sin 2\pi n f_m}{f_s}, \qquad -\infty \le n \le \infty, \qquad (11)$$

has a line spectrum given by its z-transform (Fig. 7). If the signal is truncated so that only a finite length N is nonzero, each of the ideal spectral values shown in Fig. 7 is spread, or "leaked," over the entire (continuous) spectrum (Fig. 8).

The two parts of the leaked spectrum can be called[12] a *direct contribution*, centered on f_m, and an *image contribution*, centered on $-f_m$. Figure 8 appears to show the magnitudes of the spectral contributions. Actually, each contribution is twisted helically about the frequency axis, and the complex spectrum (Fig. 9) is the frequency-dependent vector sum of the two contributions. The lack of symmetry about the frequencies $\pm f_m$ in Fig. 9 is a result of leakage from direct and image contributions into each other.

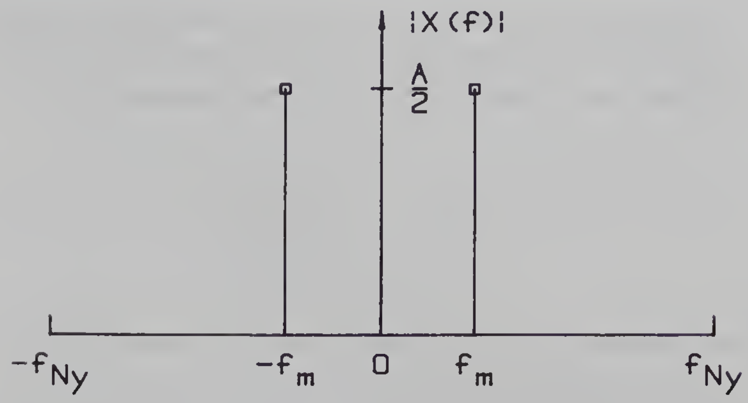


Fig. 7 Ideal spectrum of an infinite-length sinusoidal signal (z-transform); also the DFT spectrum of a sinusoidal signal that is periodic in the data record length.

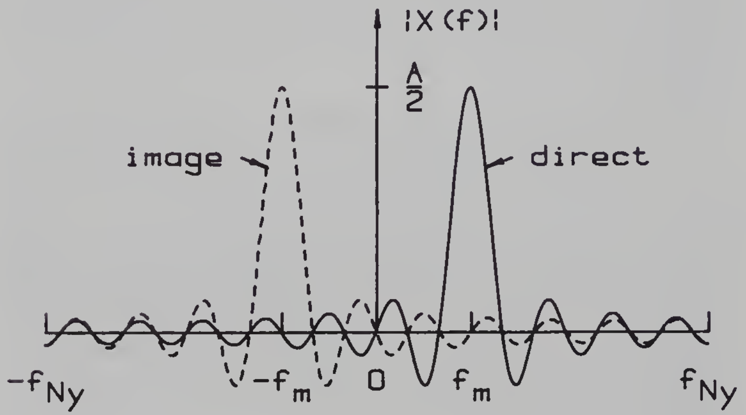


Fig. 8 Direct and image contributions to the magnitude spectrum of a finite-length sinusoidal signal (z-transform). Smearing of the spectrum is called "leakage."

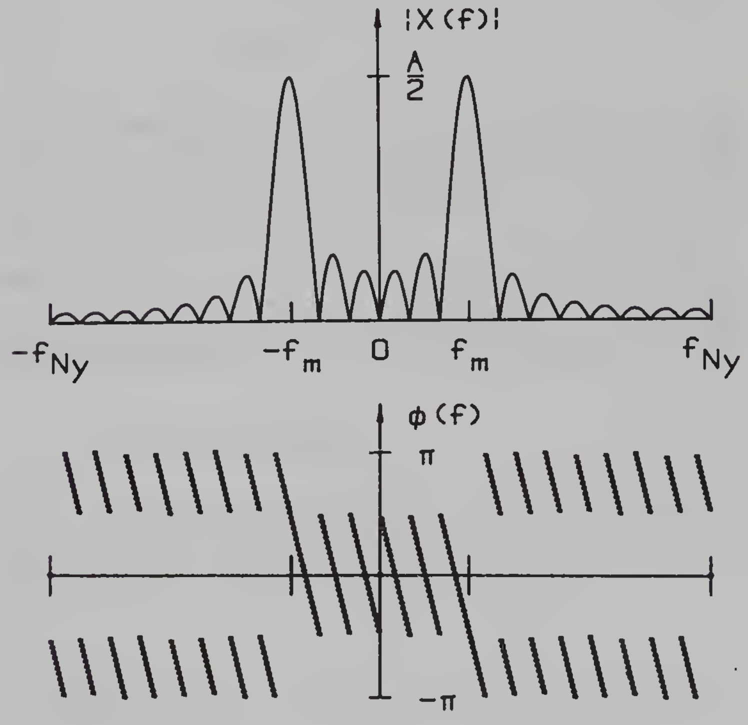


Fig. 9 Magnitude and phase spectra of a finite-length sinusoidal signal (z-transform).

Applying a z-transform to a signal truncated to finite length N is mathematically equivalent to multiplying the infinite-length signal by the weighting function

$$b(n) = \begin{cases} 1.0, & 0 \leq n \leq N-1, \\ 0, & \text{elsewhere.} \end{cases} \tag{12}$$

This function is called a *data window*. The specific data window shape given by Eq. (12) is usually called a *rectangular window* (also "gate" or "boxcar" function). As used in Eq. (12), "rectangular" has literal meaning only when used with a finite Fourier transform or a z-transform. When this window shape is used with a DFT, *open window* is more descriptive, since Eq. (12) periodically repeated looks like a constant over all time.[12] "Open" implies that everything is seen through the window that can be seen.

The spectrum of a data window is called a *spectral window*. Spectral windows typically have a *main lobe* and *side lobes* on each side of the main lobe. The main lobe of the open spectral window has a width $2/T_r$, while the side lobes have half that width (Fig. 10). Some authors use "function" or "weighting" to mean "data window" and "window" to mean "spectral window."

wnen a data window is used, the signal analyzed by a DFT is the product

$$\tilde{x}(n) = b(n)x(n). \tag{13a}$$

The spectrum of $\tilde{x}(n)$ is the convolution

$$\tilde{\mathbf{X}}(k) = \mathbf{B}(k) * \mathbf{X}(k). \tag{13b}$$

Thus the leaked spectrum of a truncated sinusoidal signal (Fig. 9) is equivalent to an ideal two-sided spectrum (Fig. 7) convolved with the spectrum of the open data window (Fig. 10).

Leakage means that the appearance of a DFT spectrum can be misleading. If a sinusoid is periodic in the data record length, the DFT spectrum looks like the true spectrum (compare Figs. 7 and 11). But leakage is not avoided; it is hidden and will appear if the spectrum is interpolated.

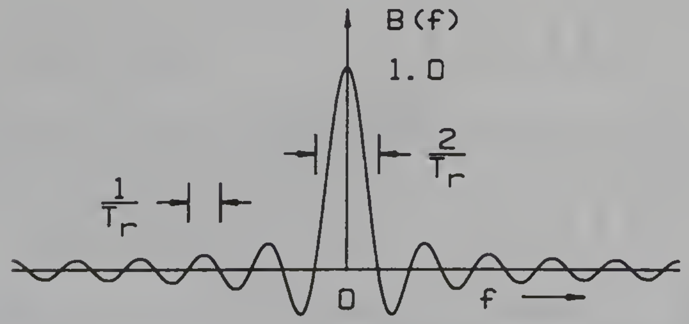


Fig. 10 Spectral window for a rectangular data window (z-transform).

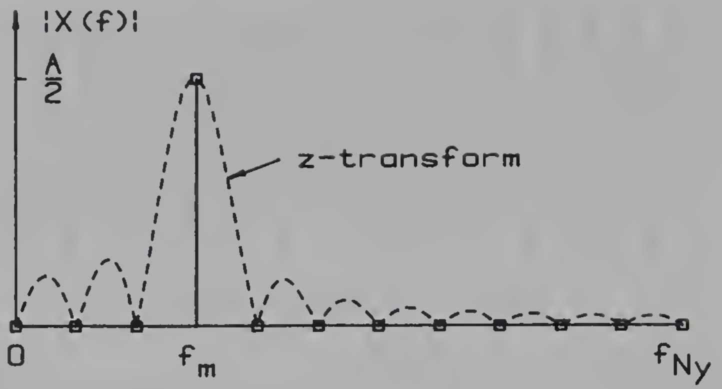


Fig. 11 Positive half of the two-sided magnitude spectrum for a data record that has exactly 3.0 periods taken from a sinusoidal signal. A z-transform (- - -) is superposed on the DFT values.

The usual case is that sinusoids are not periodic in the data record length. The worst case occurs when the data record has an extra half period of a sinusoidal signal[12] (Fig. 12). This effect is independent of N. While the main-lobe maximum is now hidden, since it corresponds to none of the DFT magnitudes, it can be computed (Section 4.3) or estimated by interpolation (Section 2.7).

Figure 12 also illustrates the possibility of a subtle error. When a signal has periodic components, connecting DFT magnitude and phase values with straight-line segments can imply an incorrect spectrum.

The misleading concept of "bins" is sometimes used to describe DFT components. The concept implies that a DFT acts like an array of narrow-band filters. While this has merit if the signal is band-limited white noise, is periodic in the data record length, or is a pulse captured entirely within the data record, it is not correct in the presence of leakage, which results in spectral spreading into all the bins. In particular, fractional octave band analysis can be corrupted by leakage.

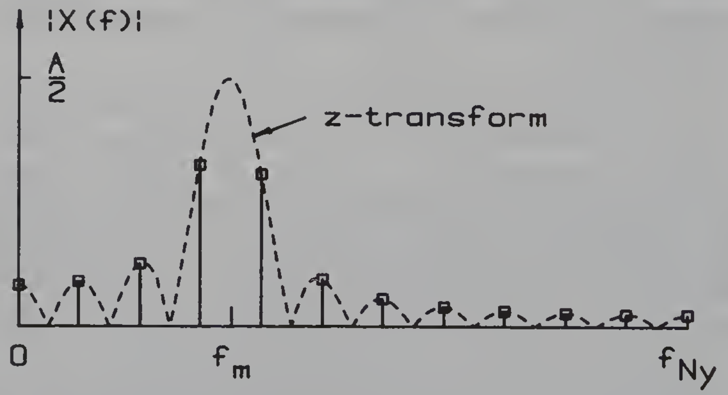


Fig. 12 Positive half of the two-sided magnitude spectrum for a data record that has 3.5 periods taken from a sinusoidal signal. A z-transform (- - -) is superposed on the DFT values.

2.7 Interpolation

Interpolation means finding values of a function between known discrete values, and it implies that a continuous function exists from which the discrete values were taken. There are two general procedures: formal interpolation and padding with zeros. Interpolation works with the data record, not the original signal. Aliasing and leakage are preserved.

Signal Interpolation A data record of size L can be interpolated to size $N = mL$, m any integer, by rewriting Eq. (1a) with Eq. (10) to get

$$x(n) = \frac{1}{N}\left\{X(0) + 2\sum_{k=1}^{L/2-1} X(k)\cos\left(\frac{2\pi kn}{N} + \phi(k)\right) + X\left(\frac{L}{2}\right)\cos\left[\frac{\pi Ln}{N} + \phi\left(\frac{L}{2}\right)\right]\right\},$$

$$0 \le n \le N. \tag{14}$$

In the last term, the argument $\pi Ln/N$ must be $\pm\pi$, and the phase angle $\phi(\frac{1}{2}L)$ is required, since it may be $\pm\pi$ (e.g., Fig. 4). Signal interpolation is also called *signal reconstruction.*[13] Only if $X(\frac{1}{2}L) = 0$ can interpolation in the time domain also be accomplished by padding with zeros in the frequency domain (evenly spaced about $\frac{1}{2}L \rightarrow \frac{1}{2}N$).

Spectrum Interpolation Calculating a DFT by using Eq. (1b) is equivalent to sampling a continuous z-transform at regular frequency intervals.[5] This means that a data record has a continuous spectrum and that details of this spectrum can be revealed by interpolating the discrete DFT spectrum. Formal interpolation requires summing a sequence of basis functions [see Ref. 6, Eq. (2.142)]. A simpler way is to pad the data record with zeros (append zero values to the end of the data record).

3 DIGITAL MEASUREMENT SYSTEM

Figure 13 represents the major components in a typical digital data acquisition and analysis system.

- A time-varying physical phenomenon of any type, $p(t)$, is sensed by a transducer assumed to have an electrical output linearly proportional to $p(t)$. With transducers having a high electrical impedance, such as condenser microphones and piezoelectric accelerometers, considerable care is required in setting up the required instrumentation to avoid problems resulting from line capacitance, stray electromagnetic noise, and ground loops.[14,15]
- Signal conditioning implies electronic devices to assure that an adequate voltage $x(t)$ is created proportional to $p(t)$ (e.g., analog-to-digital converters normally require input voltages in the range 10 mV to 10 V).
- Normally, $x(t)$ is further conditioned by an anti-aliasing filter (AAF, see Section 3.1).
- The analog signal $x(t)$ is then converted to a digital signal $x(n)$ by an analog-to-digital converter (A/D, see Section 3.2).
- The data can then be stored for later processing or used on-line (e.g., for adaptive sound control).

Every step of the data acquisition process from the physical signal $p(t)$ to the digital signal $x(n)$ is subject to error. The signal sensed by the transducer is not identical to the signal that would exist in the absence of the transducer (e.g., refraction effects at a microphone, mass loading of a structure by an accelerometer). Transducers, signal conditioning, AAFs, and A/Ds can introduce magnitude and phase distortion. Sample-and-hold (S/H) modules in the A/Ds are subject to amplitude and timing errors (e.g., jitter: small variations in the sampling interval that depend on signal amplitude[10]). Quantization and granularity errors result from the A/D process. Harmonic distortion can be introduced by any analog process, including the analog front end of an A/D converter.

Taken as a whole, these measurement errors limit the effective dynamic range it is possible to achieve and therefore limit the nominal resolution it is reasonable to specify for an A/D converter. A 16-bit normal resolution is near the optimum for most practical purposes.[10] The effective resolution can be improved by oversampling (Section 3.3).

3.1 Anti-Aliasing Filters

There is some art in selecting an AAF. Its magnitude, phase, and transient response characteristics must be considered relative to the signal being acquired.

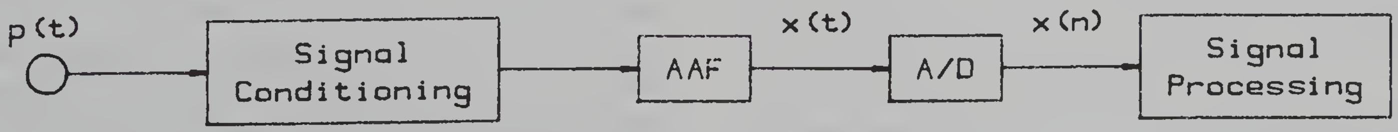


Fig. 13 Data flow in a typical digital data acquisition system.

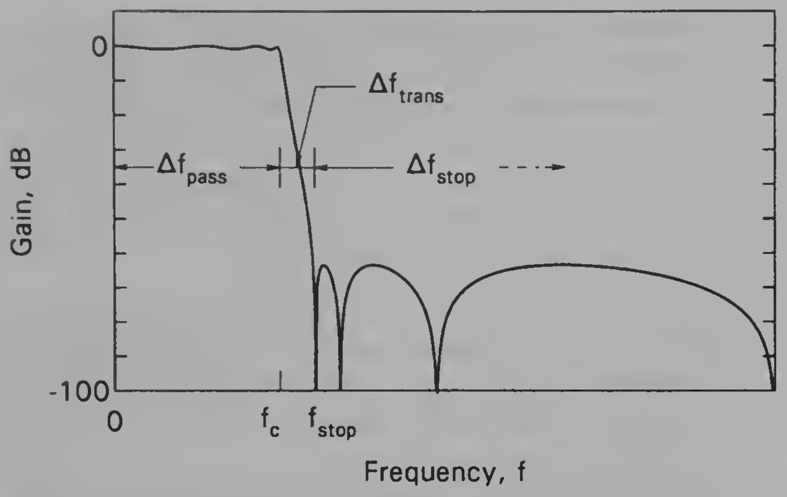


Fig. 14 Magnitude response (gain) of a typical anti-aliasing filter, showing a passband, transition band, and stopband.[13]

Magnitude Response An AAF's magnitude response is characterized by a passband Δf_{pass}, a transition region (also called "guard band") Δf_{trans}, and a stopband Δf_{stop} (Fig. 14). Both the passband and the stopband normally have a "ripple" (Figs. 14 and 15), meaning that the gain varies between upper and lower limits within the band. The upper frequency limit of the passband, f_c (cuttoff frequency), and the lower frequency limit of the stopband, f_{stop}, are defined as shown in Figs. 14 and 15. An optimum AAF should fulfill the conditions $f_{\max} \leq f_c$ and $f_s \geq 2f_{\text{stop}}$, where $f_{\max}$ is the highest frequency at which information is desired in the signal.

For use with stationary signals, flat response in the passband (usually low ripple) and rapid decay in the transition region are desirable criteria. These criteria are adequate if the application is not sensitive to phase distortion. An AAF's transient response (Fig. 16) must die out before a stationary input–output relationship can exist.

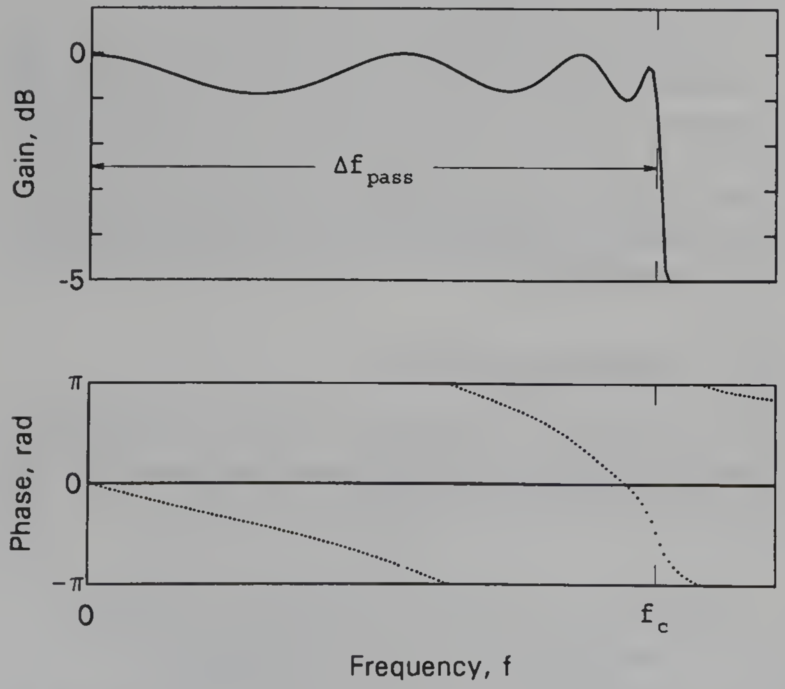


Fig. 15 Magnitude and phase response in the passband of the anti-aliasing filter of Fig. 14.

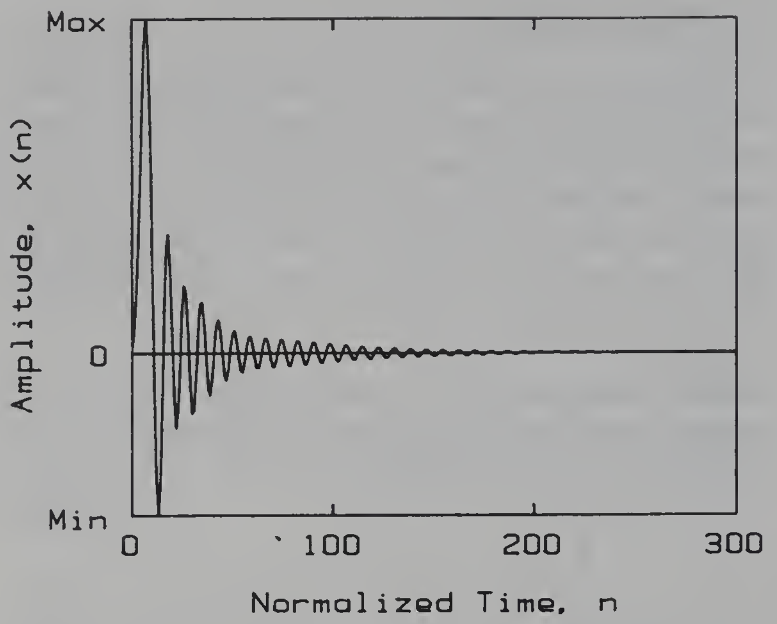


Fig. 16 Impulse response of the anti-aliasing filter of Figs. 14 and 15.

Phase Response Generally, AAFs have poor phase response, typically being nonlinear over a large part of the passband (Fig. 15). The resulting phase distortion can affect the shape of a signal. When preservation of signal shape is important, deconvolution (Section 5.3), oversampling (Section 3.3), or an all-pass filter[16] can be used to decrease phase distortion. If the signal is minimum phase, a correct signal phase can be re-created by using a Hilbert transform.[17]

Transient Response If the signal to be acquired is a transient (pulse), the AAF's output will be the convolution of the signal and the impulse response of the AAF (Fig. 16). There are three ways to decrease the effects of AAF transient response on a signal: (1) use a Bessel filter,[18] (2) deconvolve the AAF impulse response from the AAF output (Section 5.3), and (3) oversample (Section 3.3).

3.2 Analog-to-Digital Converter

An A/D converter is designed to convert an analog signal $x(t)$ into a digital signal $x(n)$ at a constant sampling rate f_s. Each discrete signal value $x(n)$ is stored in computer memory as the bits of a digital word. The signal $x(t)$ is acquired by a sample-and-hold (S/H) module, which operates under computer control to hold snapshot values of $x(t)$ long enough for the A/D converter to set all the bits it is designed to resolve and to transfer them to a digital word.

For data acquisition in acoustics, A/D converters usually have a hardware resolution of 12, 14, or 16 bits, called *nominal resolution* here. When 1 bit is used as a sign bit, *actual resolution* is reduced by 1 bit. The *dynamic range* of the digitizing process is set by quantizing noise. If the quantizing noise is uniformly distributed over the Nyquist range (band-limited white noise), the maximum dynamic range can be expressed in terms of a signal-to-noise ratio[19]:

$$(\mathrm{S/N})_{\max} \simeq 6.02(M + 0.5L) + 1.76\ \mathrm{dB}, \qquad (15)$$

where M is the actual resolution of an A/D converter in bits and L is the number (not necessarily integer) of octaves of oversampling.[19] For example, if the actual resolution is $16 - 1 = 15$ bits and there is no oversampling, $(\mathrm{S/N})_{\max} = 92$ dB. Dynamic range can be adversely affected by signal level and form (Section 3.3). It may be improved by oversampling or, if noisy, by averaging (Section 6.7).

3.3 Oversampling and Resolution Enhancement

Oversampling means using a sampling rate intentionally greater than the minimum required to prevent aliasing. As a general rule, it is advisable to operate an A/D converter at the highest possible sampling rate, so that L in Eq. (15) is large. Oversampling provides a number of practical benefits.[10,18,19] The three results most important to users are (1) AAF requirements can be relaxed, (2) dynamic range can be increased, and (3) phase response can be improved. The greater the oversampling, the less will be the need for expensive AAFs. A simple *RC* filter or even no filter can often suffice.

Oversampling increases the maximum dynamic range (i.e., effective resolution) of an A/D converter. Equation (15) implies an increase of 3 dB in $(\mathrm{S/N})_{\max}$ for each doubling of f_s beyond $(f_s)_{\min}$. The actual increase is signal dependent. The ideal signal is broadband with amplitudes that use all the bits of the converter. Doubling the sampling rate then halves the power of the quantization noise in the original Nyquist range. The ideal result is degraded if the signal is low frequency or activates only the lowest bits of the converter. Adding a dither (small-amplitude, high-frequency signal) to $x(t)$ can improve the resulting spectrum for such nonideal signals.[19] The effective S/N is decreased by 6 dB for each bit not used.

4 SPECTRAL ANALYSIS OF SIGNALS

The three classes of signal usually encountered are pulse, periodic, and random. There is no aliasing if $f_s \geq 2f_{\max}$. The two major analysis problems encountered concern leakage and noise. Spectral analysis of pulse and periodic signals is discussed in this section. Spectral analysis with noise and spectral analysis of random signals are discussed in Section 6.

4.1 Pulse Signals

A pulse is defined as a signal with a finite time duration T_p. There is no leakage if $T_r \geq T_p$. Except for the effects of noise, a DFT then gives a complete and accurate spectrum of the pulse. The DFT can be interpolated as desired to disclose the fine structure of the spectrum (Section 2.7). Useful broadband pulses for system identification include chirps[1] and a finite-length pulse taken from a white band-limited random process.

4.2 Stationary Signals, Windows, and Leakage Control

Almost always, leakage (Section 2.6) will corrupt the DFT spectrum of a stationary signal. *Leakage control* is the process by which spectral side-lobe amplitudes resulting from finite-length data records are decreased by using a shaped data window. Some authors use "tapering" to describe this process. While there seems to be a semi-infinite number of data windows,[20] very few are needed to control leakage. These few are expressed by a Fourier series,

$$b(n) = \sum_{i=0}^{M-1} B_i \cos\frac{2\pi in}{N}, \qquad 0 \leq n \leq N-1, \qquad (16)$$

where i is an integer, M is the number of (one-sided) Fourier coefficients, and N is even. Only four coefficients are needed to reduce side-lobe amplitudes by 98 dB.[21,22]

The open window is a one-coefficient window ($M = 1$) defined by $B_0 = 1$. Two well-known two-coefficient windows ($M = 2$) are the Hann (often called Hanning) and the Hamming windows (Table 4). Hamming window coefficients are often shortened to $B_0 = 0.54$, $B_1 = -0.46$.

Figures 17*a–d* show normalized spectra for several windows. They illustrate the three most important aspects of window spectra (Table 4): main-lobe width, side-lobe depression, and rate of side-lobe decay. Main-lobe width generally increases as side lobes are depressed and as side-lobe decay rate is increased. The practical advantage of selecting a data window with side lobes depressed into a signal's noise level is that the significant effects of leakage are then concentrated in a frequency band equal to the main-lobe width.

For many applications in acoustics, it is desirable to use data windows having the narrowest possible

TABLE 4 One-Sided Fourier Series Coefficients for Several Data Windows[a]

	SAR (dB)	S (k)	SLDR (dB/octate)	B_0	B_1	B_2	B_3
Open	−13.46	1.626	6	1.0			
Hann	−31.47	3.743	18	0.5	−0.5		
Hamming	−43.19	3.846	6	0.53836	−0.46164	0.0	
Opt59	−59.72	5.053	6	0.461445	−0.491728	0.046827	0.0
Opt71	−71.48	5.907	6	0.424379	−0.497340	0.078281	0.0
Opt98	−98.17	7.929	6	0.363579	−0.489179	0.136602	−0.010640

[a] Coefficients have signs that place the data window maximum at $n = \frac{1}{2}N$ and are normalized so that $\sum(-1)^i B_i = 1$. The selectivity amplitude ratio (SAR) is the ratio of the highest side-lobe to main-lobe amplitudes. Selectivity (S) is the width of the main lobe at the SAR. SLDR is the side-lobe decay rate. Values of B_i given as 0.0 are zero in fact ("most efficient" windows); those not given are unknown but taken to be zero.

Source: Adapted with permission from Ref. 21.

main lobe for a specified side-lobe depression. While Dolph–Chebyshev windows have this ideal characteristic, they have the disadvantage that N different coefficients must be computed for each value of side-lobe depression and for each data record size N.

The narrow main lobes of Dolph–Chebyshev windows can be closely approximated by optimized windows described by Eq. (16) with $M \leq 4$ coefficients (Figs. 17*c*, *d*). Fourier series coefficients B_i are given in Table 4 for several "most efficient" optimum windows.[21]

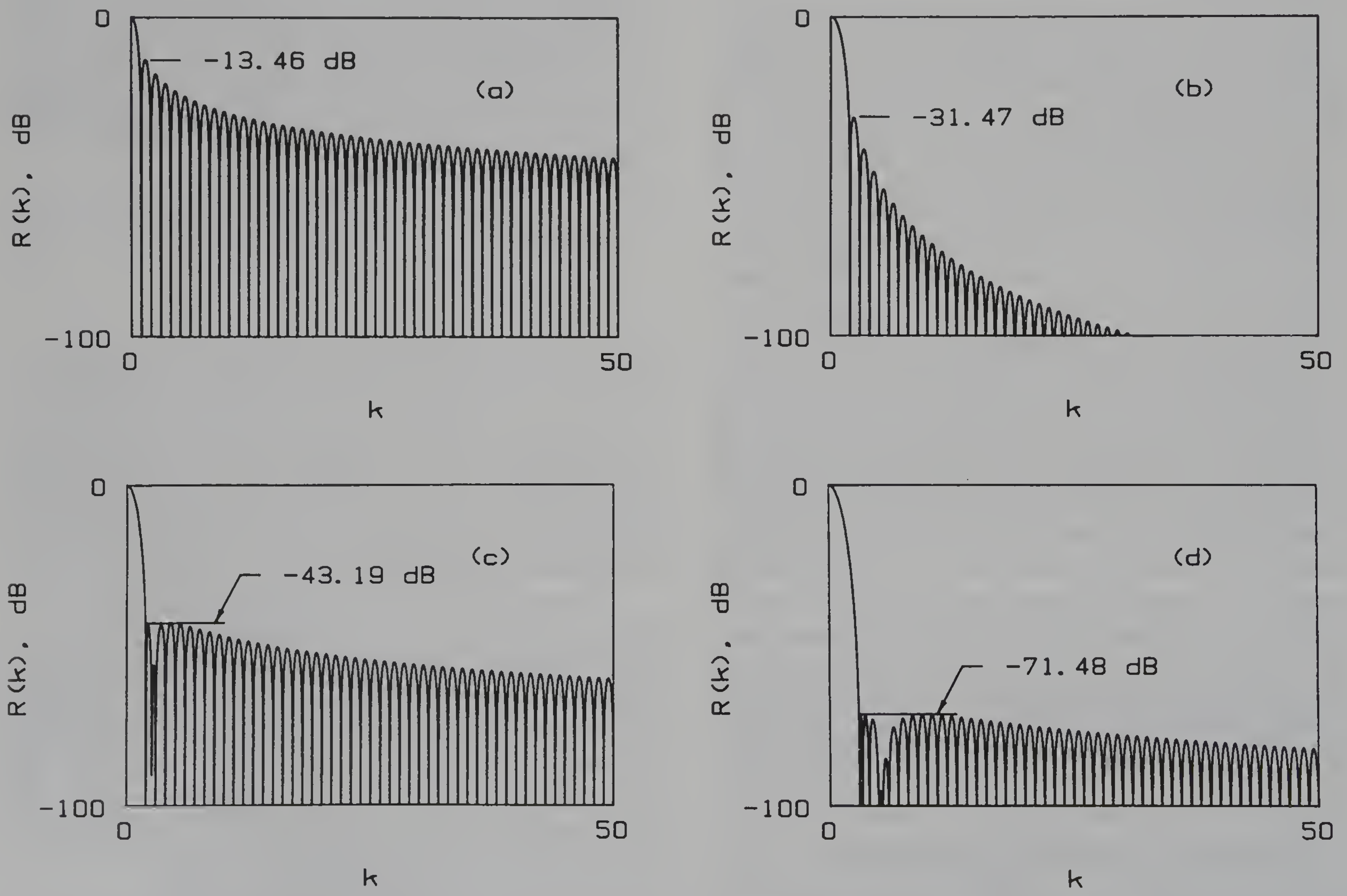


Fig. 17 Normalized spectral windows for (*a*) open (rectangular) data window, (*b*) Hann data window, (*c*) Hamming data window, and (*d*) Opt71 data window (Table 4).

4.3 Accurate Estimation of Sinusoidal Signal Parameters

It is often desirable to estimate accurate values for the amplitude, frequency, and phase of one or more sinusoids in a signal using a DFT. This requires only two complex DFT values, called *principal values*, for each sinusoid, one on each side of the frequency at which the main-lobe maximum is expected. Computation details are available.[12,23]

5 SYSTEM IDENTIFICATION

System means an analog linear, time-invariant (LTI) or discrete linear shift-invariant (LSI) operation that transforms an input signal into an output signal (Fig. 18). "System" does not refer to hardware, but to the operations performed by hardware or software.

The practical problem is to model a real LSI system with a computer-simulated LSI model using experimentally measured data. This usually means representing a distributed-parameter physical system by a lumped-parameter model.

5.1 System Identification

The operation of a discrete system (model) can be represented by the backward-difference equation

$$\sum_{i=0}^{N-1} A(i)y(n-i) = \sum_{i=0}^{M-1} B(i)x(n-i), \tag{17}$$

where $x(\cdot)$ and $y(\cdot)$ are input and output sequences indexed backward from present values $x(n)$ and $y(n)$ and M and N represent the number of coefficients (sequence lengths). The corresponding system function is[5] (z-transform)

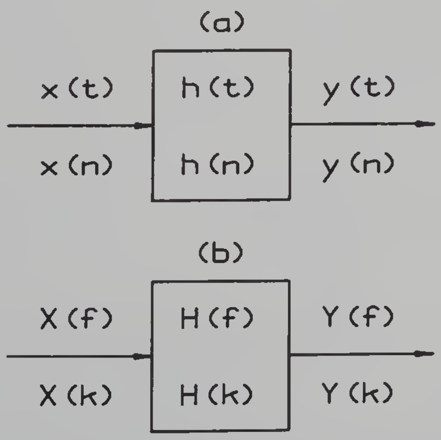


Fig. 18 Block diagrams of input/output relationships for analog (LTI) and digital (LSI) systems in the (*a*) time and (*b*) frequency domains.

$$H(z) = \frac{\sum_{i=0}^{M-1} B(i)z^{-i}}{\sum_{i=0}^{N-1} A(i)z^{-i}}, \tag{18}$$

where the system gain is $G = B(0)/A(0)$. The $M-1$ roots of the numerator polynomial are called the *zeros* of the system, and the $N-1$ roots of the denominator polynomial are called the *poles* of the system. Complex-conjugate poles define resonances, and complex-conjugate zeros define antiresonances.

System identification often means estimating numerical values for the system parameters $B(i)$ and/or $A(i)$ for use in an FIR or IIR model.[3,24,25] A system can have many parametric models, and the number of parameters and their values can be wildly different (e.g., an FIR model of a one-coefficient IIR system may need 20–100 coefficients). A system can be identified also by its impulse response, frequency response, or pole–zero values. The four ways are equivalent.

A system is *causal* if it cannot produce an output prior to receiving an input (it cannot predict). All real physical systems are causal. A causal discrete system is *stable* if all its poles are inside the unit circle. A causal, stable discrete system is *minimum phase* if all its zeros are inside the unit circle.

The (continuous) frequency response is given by evaluating Eq. (18) on the unit circle, $z = e^{j\theta}$, $\theta = 2\pi k/N$, $k/N = f_{\text{norm}}$, where N can be selected to provide any Δf. If N is selected for an FFT, fast deconvolution (Section 5.3) and interpolation can be used.

5.2 Convolution

Convolution is the process by which an LTI or LSI system transforms an input $x(\cdot)$ into an output $y(\cdot)$ (Fig. 18). For an LSI system initially at rest or after transient response has disappeared, convolution is defined by the sequence

$$y(n) = \sum_{i=0}^{n} x(i)h(n-i), \qquad n \geq 0, \tag{19}$$

or, equivalently, by the *response function*

$$\mathbf{Y}(k) = \mathbf{H}(k)\mathbf{X}(k). \tag{20}$$

When a DFT is used, Eq. (20) results in a circular response function $\mathbf{Y}^c(k)$. A circular convolution sequence is given by

$$y^c(n) = \text{IDFT}\{\mathbf{Y}^c(k)\}. \qquad (21)$$

The process is called *fast convolution* when an FFT is used.

When signals are applied to physical systems, the process is called *linear convolution*. While circular convolution does not usually give the same results as linear convolution, a linear convolution can often be recovered (Section 5.4).

5.3 Deconvolution

Deconvolution means undoing the effects of convolution. It can be used to estimate $x(n)$ or $h(n)$. Both are implemented in the frequency domain by rearranging Eq. (20). Deconvolution can be implemented also by cepstral analysis (Ref. 5, Chap. 12). The effects of noise can be reduced (Section 6.7).

System Input Signal Estimation A system's input can be estimated when its output and impulse responses are known. From Eq. (20),

$$\mathbf{X}(k) = \frac{\mathbf{Y}(k)}{\mathbf{H}(k)}, \qquad (21a)$$

$$x(n) = \text{IDFT}\{\mathbf{X}(k)\}. \qquad (21b)$$

Deconvolution, in seeking an input given an output, amounts to prediction, a noncausal process, and it is therefore impossible to implement with analog systems.

System Impulse Response Estimation A system's frequency response and impulse response can be expressed by

$$\mathbf{H}(k) = \frac{\mathbf{Y}(k)}{\mathbf{X}(k)}, \qquad (22a)$$

$$h(n) = \text{IDFT}\{\mathbf{H}(k)\}. \qquad (22b)$$

Equations (22) are most useful with broadband deterministic signals and low measurement system noise levels. Chirps and choips[1] are particularly useful, although care must be taken to avoid an excitation data record (not necessarily the true signal) with mean value zero or near to it. A DC bias can always be added.

Phase computed with a DFT may not be unique, since DFT phase has only the range $-\pi \le \phi \le \pi$. For example, a true phase 2.5π will be computed as 0.5π. Recovering the true phase is called *phase unwrapping*.[5,26]

Deconvolution involves inverse filtering.[17] If a filter has any finite zeros, they become poles of its inverse, $H_{\text{if}}(k) = 1/H(k)$. The practical result is that a data record length N satisfactory for $H(k)$ may not be long enough to adequately represent $H_{\text{if}}(k)$. Errors can result if the inverse filter has small spectral values, especially if they are affected by noise.

5.4 Linear Convolution and Correlation

Special care is required to realize a linear convolution or correlation when using the intrinsically circular DFT operation. The general procedure is to assure that at least one of the two sequences has an adequate number of zero values.

If the sequences are pulses of lengths L and P, data record lengths $N \ge L + P - 1 + D$ (Fig. 19) will result in a linear realization, where D represents output delay relative to input.

With continuing waveforms, truncation and zero padding are required.

A typical objective is to convolve a long sequence with a short sequence (Fig. 20). For example, $h(n)$ may represent an impulse response of length P. If a data record $x(n)$ can be taken with length $N > P$, a linear convolution $y(n)$ of length $N - P$ can be realized by setting to zero the last P values of $x(n)$ to get $x'(n)$, then convolving $x'(n)$ with $h(n)$. When the original $x(n)$ cannot be entirely contained within a single record length,

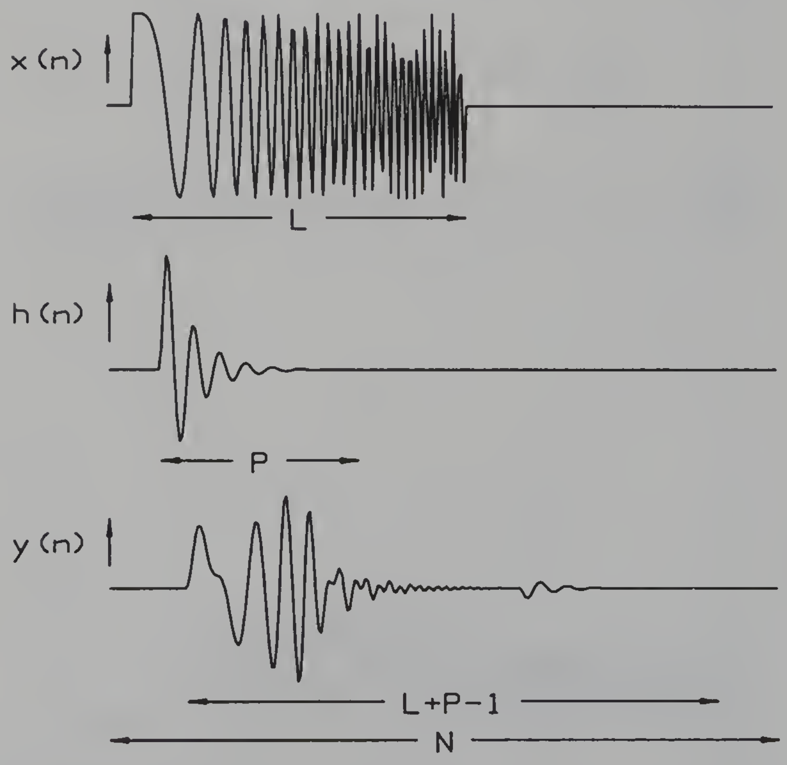


Fig. 19 Length relationships among a pulse excitation $x(n)$ (chirp), a system impulse response $h(n)$, and a system response $y(n)$ to allow linear convolution/deconvolution to be recovered from circular convolution/deconvolution.

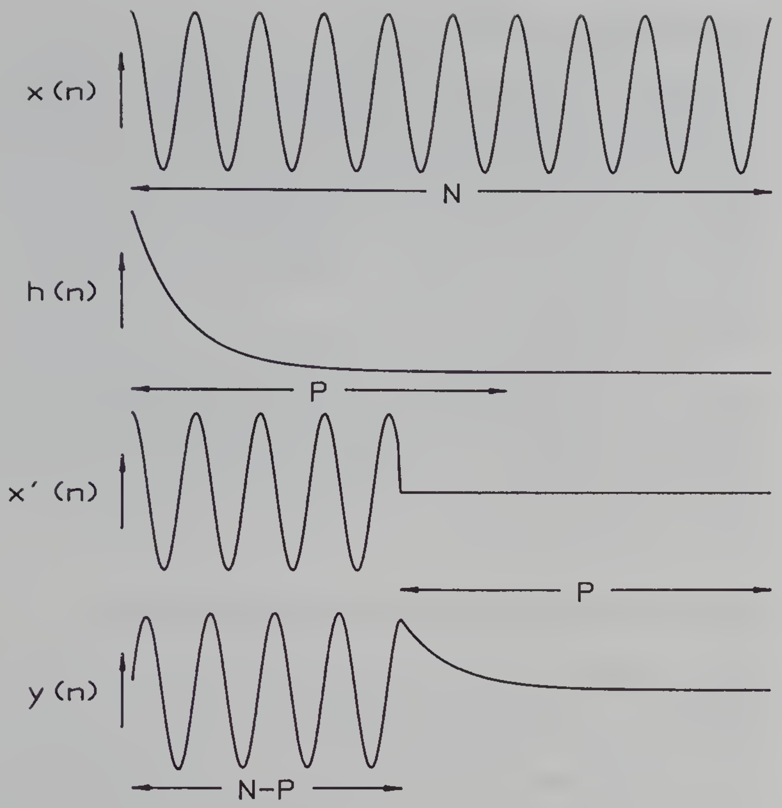


Fig. 20 Length relationships between a data record from an extended excitation $x(n)$ and a system impulse response $h(n)$ to allow recovery of a linear convolution $y(n)$ from circular convolution.

an extended $y(n)$ can be found using overlap-save or overlap-add methods.[5,6]

Another typical objective is to find the linear autocorrelation sequence $r_{xx}(\tau)$ of a signal $x(n)$ (Fig. 21). The procedure is to create a signal $y(n)$ by setting to zero the last P values of $x(n)$, calculate the energy spectrum $\mathbf{S}^c_{xy}(k) = \mathbf{X}^{c*}(k)\mathbf{Y}^c(k)$ (or power spectrum, Section 6.3), and take the IDFT to get $r^c_{xy}(\tau)$. The first P values of $r^c_{xy}(\tau)$ are identical to those of the linear $r_{xx}(\tau)$. Since an autocorrelation sequence is even $[r_{xx}(-\tau) = r_{xx}(\tau)]$, the result can be used to get a linear autocorrelation for $|\tau| \leq P$. A linear cross-correlation sequence can be calculated in a similar way, with care taken to calculate separately for $\tau < 0$ and $\tau > 0$, since cross-correlation sequences usually are not even. When $x(n)$ is a random signal, $P \geq \tau_{cl}$ and averaging are necessary (Section 6).

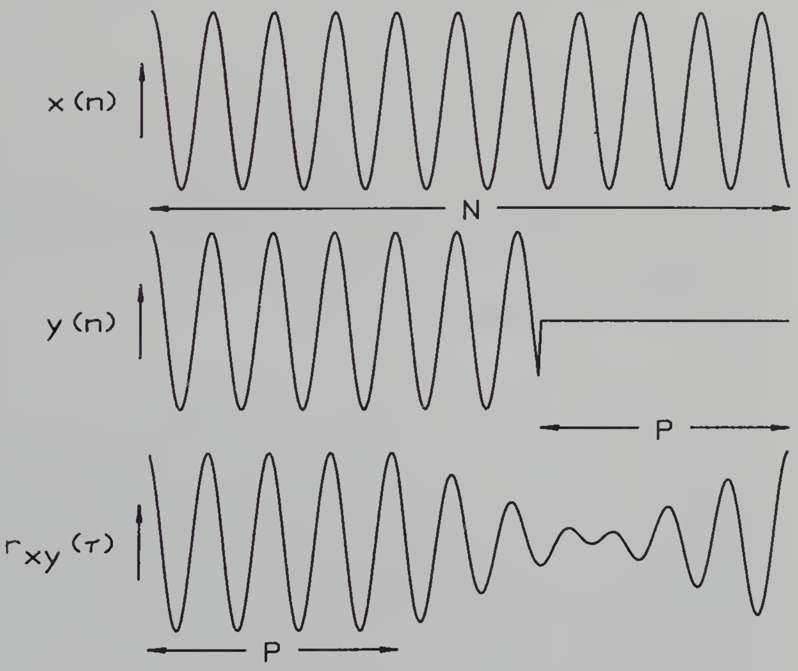


Fig. 21 Length relationships for a data record to allow recovery of a linear correlation from a circular correlation.

6 NOISE AND RANDOM SIGNALS

Analysis objectives for random-signal components include reducing the effects of unwanted noise, finding the frequency ranges where signal power is significant, and identifying systems. The practical problem is to achieve analysis objectives using a DFT, a method that treats each data record as if the underlying process were periodic.

Random-signal analysis requires working with bias and random errors. Emphasis in this section is placed on limiting the effects of random error. The procedures discussed are based on first- and second-order statistics (mean, variance, correlation), stationary random processes, and single-input, single-output (SISO) systems. Higher order statistics,[27] nonstationary signals,[28–30] and multiple input–output systems[9] are discussed elsewhere. Much information is available concerning random-signal analysis.[2,3,9,13,31,32]

6.1 Ideal Random Signal

An *ideal* random signal $x(n)$ is defined here as stationary, with a band-limited white spectrum, a normal amplitude distribution, and an infinite duration. A major effect of this assumption is to provide a base measure for the amount of random data to acquire.

Random signal means that $x(n)$ is a single realization, $-\infty \leq n \leq \infty$, from a *random process* (also called *stochastic process*).

White means that $x(n)$ has a constant-magnitude power spectrum. It means also that $x(n)$ is an *independent* random variable, that no value of the sequence is influenced by any other. *Band limited* means that $x(n)$ has power only in the Nyquist range. When practical random signals do not originate in a white process, they can often be regarded as the result of passing a white random signal through a linear filter. *Practical* also means finite-length data records and sample spectra. Sample spectra taken from a white process will not be white, and spurious spectral peaks and valleys can be nasty sources of interpretational error (Ref. 2, Chap. 7).

Stationary means that the statistical moments of the process do not change with time. For practical purposes, a stationary random process may be assumed to be *ergodic*, meaning that its ensemble moments and time moments have the same descriptive properties. *Weakly stationary* means that only the first and second moments (mean, variance) are stationary. *Strongly stationary* means that all moments are stationary. Practical random signals are not stationary, but are often nearly so over finite time intervals.

Normal means that a normal (Gaussian) probability density function (pdf) describes the amplitude distribution of $x(n)$. The mean, variance, standard deviation, and pdf of $x(n)$ are represented by μ, σ^2, σ, and $p(x)$ (Fig. 22*a*). Conventional practice is to replace $x(n)$, by a *standardized normal random variable*

$$z = x - \frac{\mu}{\sigma}, \tag{23}$$

where z has a zero mean, unit variance, and pdf $p(z)$. Tables of values for $p(z)$ are widely available (e.g., see Ref. 33). While second-order analysis is adequate for normal processes, most real-world signals are not normal. Much additional information can be extracted by using higher order moments.[27]

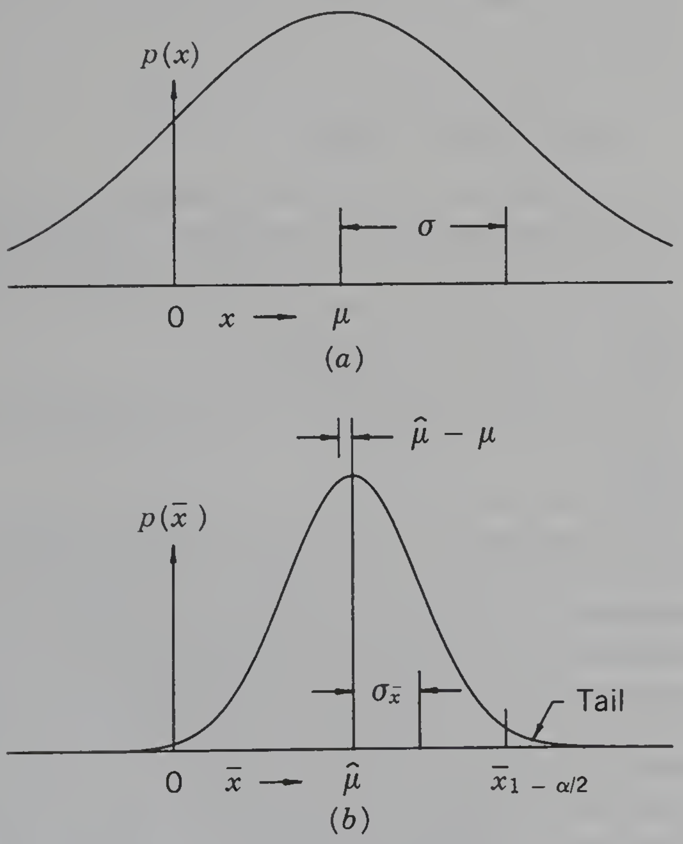


Fig. 22 Normal probability density functions: (*a*) $p(x)$ for a random variable x; (*b*) $p(\bar{x})$ for the sampling distribution of the mean.

6.2 Degrees of Freedom

Averaging is the key to achieving (statistical) *stability*, meaning the relative believability (error or precision) of an estimate. Whether random signals are averaged in the time or frequency domains, it is essential to acquire data having a sufficient number of (statistical) *degrees of freedom* ν, meaning the number of statistically independent random variables.

When enough data can be acquired to assure $\nu \gtrsim 30$, *large-sampling theory*[31] can be used in planning for data acquisition. Otherwise, *small- (exact) sampling theory* may be necessary. The emphasis in this section is on large-sampling theory.

6.3 Correlation and the Power Spectrum

Random signals can be represented by their correlation sequences or their power spectra. Normally, a power autocorrelation sequence dies out with increasing lag time τ [τ is defined in Eqs. (24)]. A useful parameter is the *correlation length* τ_{cl}, defined as the value of τ for which $r(\tau)$ decays to ϵ, an arbitrarily small quantity (e.g., system noise level).

Linear *power autocorrelation* and *power cross-correlation* sequences can be defined by

$$r_{xx}(\tau) = \frac{1}{L - \tau_{cl} + 1} \sum_{l=0}^{L-\tau_{cl}} x(l)x(l+\tau),$$
$$0 \le \tau \le \tau_{cl} - 1, \tag{24a}$$

$$r_{xy}(\tau) = \frac{1}{L - \tau_{cl} + 1} \sum_{l=0}^{L-\tau_{cl}} x(l)y(l+\tau),$$
$$0 \le \tau \le \tau_{cl} - 1, \tag{24b}$$

where the amount of data required is at least $L \ge \tau_{cl}$ and $L \gg \tau_{cl}$ represents averaging necessary to improve stability. Equations (24) provide equal stability over the entire range of τ.

Circular *discrete power spectra*

$$\mathbf{S}^c_{xx}(k) = \frac{1}{N}\,\mathbf{X}^*(k)\mathbf{X}(k) = \frac{1}{N}\,|\mathbf{X}(k)|^2,$$
$$-\frac{N}{2} + 1 \le k \le \frac{N}{2}, \tag{25a}$$

$$\mathbf{S}^c_{xy}(k) = \frac{1}{N}\,\mathbf{X}^*(k)\mathbf{Y}(k),$$
$$-\frac{N}{2} + 1 \le k \le \frac{N}{2} \tag{25b}$$

are calculated using DFTs, where the asterisk means complex conjugate. These spectra, computed from single data records $x(\cdot)$ and $y(\cdot)$, are called *periodograms*. In effect, a periodogram is a *sample spectrum*. When the Fourier components $\mathbf{S}^c(k)$ are independent random variables (ideal random process), each has only $\nu = 2$ degrees of freedom for $1 \le |k| \le \frac{1}{2}N - 1$ and $\nu = 1$ degree of freedom for $|k| = 0, \frac{1}{2}N$. The stability of a periodogram is notoriously poor, its standard deviation being equal to its mean, regardless of size N [Eq. (32)].

Average circular *power density spectra* can be estimated by dividing the discrete spectral values by the frequency interval, $\mathbf{S}^c(k)/\Delta f$.

Circular *power correlation* sequences

$$r^c_{xx}(\tau) = \mathrm{IDFT}\{\mathbf{S}^c_{xx}(k)\}, \tag{26a}$$

$$r^c_{xy}(\tau) = \mathrm{IDFT}\{\mathbf{S}^c_{xy}(k)\} \tag{26b}$$

can be calculated from Eqs. (25). Linear sequences can be recovered (Section 5.4).

6.4 Block Averaging: Sample Mean, Variance, and Spectrum

When a DFT is used, analysis objectives for stability (i.e., degrees of freedom, confidence[14]) can be accomplished by *block averaging*, meaning to average across M data records or spectra, each of size N.

Time-Domain Averaging Averaging in the time domain is done over the M block values $x_m(n)$. The sequences of N unbiased and consistent estimators are

$$\bar{x}(n) = \frac{1}{M}\sum_{m=0}^{M-1} x_m(n), \qquad 0 \le n \le N-1, \tag{27a}$$

$$s^2(n) = \frac{1}{M-1}\sum_{m=0}^{M-1} [x_m(n) - \bar{x}(n)]^2, \qquad 0 \le n \le N-1, \tag{27b}$$

where the N *sample means* $\bar{x}(n)$ are estimates of μ and the N *sample variances* $s^2(n)$ are estimates of $\sigma^2(n)$.

Data values in the sequence $x_m(n)$ are usually at least N data values apart in the sequence $x(n)$, and the sequence $x_m(n)$ will be uncorrelated if the correlation length $\tau_{cl} < N$. If the effects of correlation are not important, the sequences $\bar{x}(n)$ and $s^2(n)$ can be further averaged over N to obtain sample mean values $\bar{x}$ and s^2 over $M \times N$ data values.

Power-Domain Averaging Averaging in the power domain means averaging sample spectra. This results in a sequence of N averaged estimates of spectral values,

$$\overline{\mathbf{S}}(k) = \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{S}_m(k), \qquad 0 \le k \le N-1, \tag{28}$$

where $\mathbf{S}_m(k)$ is the kth spectral component from the mth periodogram and the $\overline{\mathbf{S}}(k)$ are sample spectral values (sample variances) [the true spectrum would be represented by $\mathbf{S}(k)$]. Since each spectral component $\mathbf{S}_m(k)$ is influenced by all values $x(n)$ in a data record, the sequence $\mathbf{S}_m(k)$ will be uncorrelated only if the M data records are separated by at least τ_{cl} data values. Even if $x(n)$ is taken from a white process, the practical restriction to a finite amount of data means that the sequence $\mathbf{S}_m(k)$ will have a random correlation.

When the variance of a random variable ϕ [e.g., $\overline{\mathbf{S}}(k)$] can be calculated, a useful relationship is the *normalized variance* (also called *dimensionless variability*[32] and *normalized mean-square random error*[9]), defined by

$$\mathrm{Var}_{\mathrm{norm}}\{\phi\} = \frac{\mathrm{Var}\{\phi\}}{[\mathrm{Ave}\{\phi\}]^2}, \tag{29}$$

where Var means "variance" and Ave means "mean value."

6.5 Some Sampling Statistics

Sampling Distribution of the Mean Sample means $\bar{x}(n)$ [Eq. (27a)] have a pdf $p(\bar{x})$ called the *sampling distribution of the mean* (Fig. 22*b*). The central limit theorem assures that $p(\bar{x})$ is very nearly normal for $\nu \gtrsim 30$ whether the pdf $p(x)$ is normal or not. When $p(x)$ is also normal, $p(\bar{x})$ can be nearly normal for $\nu < 30$, possibly[9] for $\nu \gtrsim 4$. The mean, variance, and standard deviation of $p(\bar{x})$ are represented here by $\hat{\mu}$, $\sigma^2_{\bar{x}}$, and $\sigma_{\bar{x}}$.

When both $p(\bar{x})$ and $p(x)$ are normal, the relationship between the variances of $\bar{x}$ and x is

$$\frac{\sigma^2_{\bar{x}}}{\sigma^2} = \frac{1}{\nu}. \tag{30}$$

Sampling Distribution of the Variance and the Spectrum Sample variances s^2 and $\overline{\mathbf{S}}(k)$ [Eqs. (27b) and (28)] have a pdf $p(s^2)$ called the *sampling distribution of the variance* (Fig. 23). This pdf has a chi-squared distribution with ν degrees of freedom, χ^2_ν. The mean and variance of a χ^2_ν distribution are

$$\mu_\chi = \nu, \tag{31a}$$

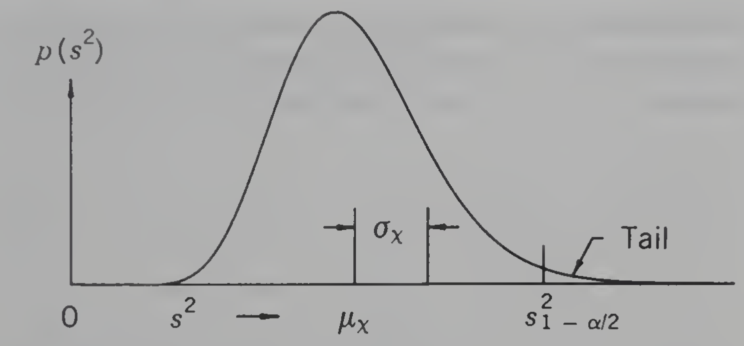


Fig. 23 Sampling distribution of the variance; chi-squared distribution for $\nu = 30$ degrees of freedom.

$$\sigma_X^2 = 2\nu. \tag{31b}$$

The ratio of the chi-square standard deviation to the mean,

$$\frac{\sigma_X}{\mu_X} = \left[\frac{2}{\nu}\right]^{1/2}, \tag{32}$$

is often used as a measure of stability.

When $p(x)$ is normal, the relationship between the sample and true variances is

$$\frac{s^2}{\sigma_X^2} = \frac{\chi_\nu^2}{\nu}. \tag{33}$$

The pdf $p(s^2)$ is very nearly normal[33] for $\nu \gtrsim 30$ (Fig. 23), and

$$\chi_\nu^2 \simeq \tfrac{1}{2}[z + (2\nu - 1)^{1/2}]^2 \tag{34}$$

can be used to calculate χ_ν^2 when a value for z (standardized normal random variable) can be specified. For example, if $z = 1.645$ (corresponds to 90% confidence[14]) and $\nu = 30$, $\chi_{30}^2 \simeq 43.5$ (true value is 43.8).

6.6 Averaging over Frequency: Windows

Spectral resolution for a random process is influenced by a data window (Sections 2.6 and 4.2), and both resolution and stability are influenced by a lag window. *Lag windows* are time-domain weighting functions similar to data windows except that they are applied to correlation sequences rather than to data records.

With a random process, using any formed window is equivalent to performing a moving average over frequency. A time-domain window (data or lag) creates a *spectral window* $W(k)$ that is convolved with a sample spectrum $\overline{S}(k)$. The process is often called *smoothing*. While the effect is similar to that of a constant-bandwidth analog swept-frequency filter, the results are not identical. When a DFT is used, the result is a circular moving average, and values near zero and Nyquist frequencies will be "aliased."

Because a spectral window acts like a swept-frequency filter, it is common practice to define *resolution bandwidth* W_e as the effective bandwidth of the "filter"[9] (also called *statistical bandwidth*).

Resolution bandwidth is determined by the data window Fourier coefficients B_i and/or the lag window coefficients W_i. It can be defined in terms of normalized variance [Eq. (29)] as (see Ref. 32, Sections 8 and 9)

$$W_e = \frac{1}{\mathrm{Var}_{\mathrm{norm}}} = \frac{[\mathrm{Ave}\{\overline{S}(k)\}]^2}{\mathrm{Var}\{\overline{S}(k)\}}. \tag{35a}$$

If the power spectrum has no sharp magnitude changes, resolution bandwidth can be approximated by (see Ref. 2, Section 6.4, and Ref. 32, Section 9)

$$W_e = \frac{[\sum B_i^2]^2}{\sum B_i^4} \tag{35b}$$

for a data window and by

$$W_e = \frac{[\sum W_i]^2}{\sum W_i^2} \tag{35c}$$

for a full-length lag window. As used here, bandwidth is defined in units of k, the frequency index. When no formed windows are used (the open-window case), $W_e = 1$ is the base resolution bandwidth set by the data record length. Multiplying W_e by $\Delta f = 1/T_r$ [Eq. (7)] gives the bandwidth in hertz.

Resolution bandwidths created by data windows are narrow, while those created by lag windows can be quite broad. Resolution bandwidths greater than the widths of peaks and valleys in a true spectrum will bias resonances negatively and antiresonances positively.

When enough data cannot be acquired to achieve a desired stability by block averaging, smoothing with a lag window is the only way to improve stability. The effect of using a window with random data is quite distinct from its effect with deterministic data (Sections 2.6 and 4.2). Data windows control leakage with both random and deterministic data. While lag windows normally do not control leakage, a unique procedure may do so.[34]

Note that the values of W_i for data and lag windows are not the same.

Smoothing with a Data Window The (power) spectral window coefficients corresponding to a data window are

$$W_i = B_i^2. \tag{36}$$

For example, the Fourier coefficients for the two-sided Hann data window $B_i = \pm\frac{1}{4}, \frac{1}{2}, \pm\frac{1}{4}$ correspond to spectral window coefficients $W_i = \frac{1}{16}, \frac{1}{4}, \frac{1}{16}$.

Unless the coefficients B_i are normalized so that $\sum W_i = 1$, the power spectrum will be biased. This normalization is different from that used for leakage control of deterministic signals (Table 4). For example, if a Hann data window is used, bias can be removed by multiplying the power spectrum by the scaling factor $1/\sum B_i^2 = \frac{8}{3}$. Using Eq. (35b), the resolution bandwidth for the Hann data window is $W_e = 2$.

The different normalizations required for leakage control and smoothing result in different calibration factors for periodic and random components in a signal[12].

A peculiar characteristic of data windows is that they increase stability by averaging over frequency, but they lose the same amount of stability by discarding useful data.[35] (The "lost" stability can be recovered by increasing the number of data records and overlapping them.[34]) Thus the degrees of freedom remain unchanged when a data window is used,

$$\nu = \begin{cases} 2 \cdot M, & 1 \le k \le (N/2) - 1, \\ M, & k = 0,\ N/2, \end{cases} \tag{37}$$

where M is the number of periodograms averaged.

Smoothing with a Lag Window The procedure is to calculate correlation sequences [Eqs. (24) or (26)], multiply by a lag window, and apply a DFT to get a smoothed spectrum.

A lag window can be represented by Eq. (16), with b, B, n, and N replaced by w, W, τ, and L, where the signs of W_i must make $w(\tau)$ an even function of τ. For a full-length lag window, $L = N$. Normally,[2,32] $L \ll N$, with $w(\tau) = 0$, $|\tau| > \frac{1}{2}L$.

The resolution bandwidth for a full-length lag window is given by Eq. (35c). For example, the Fourier coefficients for the two-sided Hann lag window of length $L = N$ are $W_i = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, and its resolution bandwidth is $\frac{8}{3}$. The signs (all positive) place the window maximum at $\tau = 0$. Bias is avoided by using the normalization $\sum W_i = 1$.

Increasing the resolution bandwidth corresponds to increasing the number of degrees of freedom to

$$\nu = \begin{cases} 2 \cdot M \cdot W_e, & 1 \le k \le (N/2) - 1, \\ M \cdot W_e, & k = 0,\ N/2. \end{cases} \tag{38a}$$

When lag window length $L \ll N$, a conservative approximation for the degrees of freedom is given by[32]

$$\nu \simeq \begin{cases} \dfrac{M \cdot N}{L}, & 1 \le k \le (N/2) - 1, \\[2ex] \dfrac{M \cdot N}{2L}, & k = 0,\ N/2. \end{cases} \tag{38b}$$

6.7 Noise Added to a Deterministic Signal

If a deterministic signal $x(n)$ can be repeated exactly (e.g., pulse, choip[1]), its average over M data records will not change. When the signal is corrupted by random noise $n(n)$, the result

$$y(n) = x(n) + n(n) \tag{39}$$

exhibits random amplitude variations. The effects of $n(n)$ can be reduced by averaging in either the time or frequency domain.

Time-Domain Averaging A block average over M data records in the time domain is

$$\bar{y}(n) = x(n) + \frac{1}{M} \sum_{m=0}^{M-1} n(n), \qquad 0 \le n \le N - 1. \tag{40}$$

If $n(n)$ has a nonzero mean, $\bar{y}(n)$ will be displaced from $x(n)$ approximately by this value. Otherwise, $\bar{y}(n) \rightarrow x(n)$. An advantage of this method is that results do not depend on the spectrum of the noise. Also, $\bar{n}(n)$ can be further averaged over N.

Averaging improves the *signal-to noise-ratio*, S/N, defined as the ratio of powers in signal and noise. Equation (30) can be used to estimate an enhanced S/N,

$$\mathrm{S/N} = 10 \log_{10} M + (\mathrm{S/N})_{\mathrm{orig}} \quad \mathrm{dB}. \tag{41}$$

Frequency-Domain Averaging Block averaging in the frequency domain results in

$$\overline{\mathbf{Y}}(k) = \mathbf{X}(k) + \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{N}(k), \qquad 0 \le k \le N - 1, \tag{42}$$

where $\mathbf{X}(k)$ and $\mathbf{N}(k)$ represent circular or linear spectra of deterministic and random components in $y(n)$. A practical difficulty is that the noise may not be white, and $\overline{\mathbf{Y}}(k)$ will have a frequency-dependent bias.

The spectral estimate $\overline{\mathbf{Y}}(k)$ can be improved if $\overline{\mathbf{N}}(k)$ can be separately estimated. This requires careful use of an A/D converter to assure that a noise estimate is not adversely affected by quantization and granular noise (Section 3) or by different electrical impedances in the measurement system with and without the signal source impedance. When two signals are required (e.g., deconvolution, frequency response), the two noise signals may be affected differently.

Averaging can also be done in the power domain, but phase information for $\overline{\mathbf{X}}(k)$ is lost. If $x(n)$ is minimum phase, the phase spectrum can be estimated with a discrete Hilbert transform.[17]

6.8 System Identification

Frequency response can be more reliably estimated by exciting a system with a deterministic signal than with a random signal (Section 5). If random-signal excitation is necessary, well-established procedures are available.[9,13,32]

The system to be analyzed can be represented by Fig. 24, where $n(n)$ represents noise added to the unknown true system $H(z)$ anywhere between input $x(n)$ and measured output $y(n)$, $\tilde{H}(z)$ represents an FIR model, and $\epsilon(n) = y(n) - \tilde{y}(n)$ is the residual. Minimizing $\epsilon(n)$ in the least mean-squares sense is a classic optimization problem.[24] The "optimum" frequency response is estimated by

$$\hat{H}(k) = \frac{\overline{S}_{xy}(k)}{\overline{S}_{xx}(k)}, \tag{43}$$

where the $\overline{\mathbf{S}}(\cdot)$ represent block average sample spectra [Eq. (28)]. Care must be taken to account for the effects of noise and possible small values in the inverse filter, $1/\overline{\mathbf{S}}_{xx}(k)$ (Section 5.3).

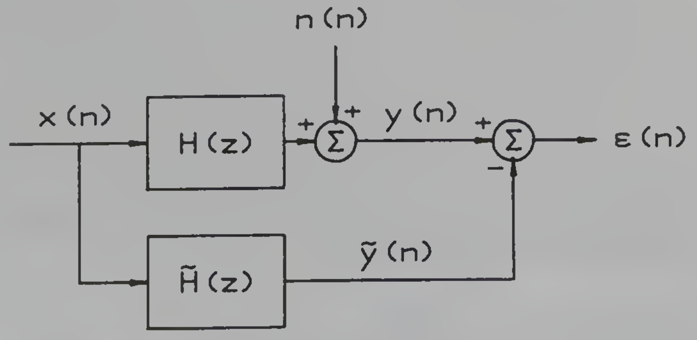


Fig. 24 Block diagram representing system identification with additive noise $n(n)$.

Equation (43) results from the optimization process whether or not there is additive noise $n(n)$ [additive noise is hidden in $\overline{\mathbf{S}}_{xy}(k)$]. The practical problem is how much of $\hat{H}(k)$ to believe. The *coherency-squared spectrum*[2]

$$\gamma_{xy}^2(k) = \frac{|\overline{\mathbf{S}}_{xy}(k)|^2}{\overline{\mathbf{S}}_{xx}(k)\overline{\mathbf{S}}_{yy}(k)}, \qquad 0 \le \gamma_{xy}^2 \le 1.0, \tag{44}$$

1.0 means that there is no additive noise; $\gamma_{xy}^2(k) = 0$ means that $\hat{H}(k)$ is all noise. The term $\gamma_{xy}^2(k)$ treats nonlinear response as if it were noise, since it is interpreted as a degradation of perfect linear response. When $\gamma_{xy}^2(k)$ shows degradation at resonant peaks, nonlinear response may be the cause. For a single data record, $\gamma_{xy}^2(k) = 1$ for all k, since a single data record is treated as deterministic by the DFT. A coherency-squared spectrum requires block-averaged spectra.

REFERENCES

1. J. C. Burgess, "Chirp Design for Acoustical System Identification," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 1525–1530.
2. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, 1969.
3. D. G. Childers (Ed.), *Modern Spectral Analysis*, Institute of Electrical and Electronics Engineers, New York, 1978.
4. J. C. Burgess, "Digital Spectral Analysis with Application to Identification of Mechanical Systems and Planning for Measurement," unpublished class notes, University of Hawaii, 1987.
5. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
6. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
7. R. N. Bracewell, *The Fast Fourier Transform and Its Applications*, 2nd rev. ed., McGraw-Hill, New York, 1986.
8. Digital Signal Processing Committee (Ed.), *Programs for Digital Signal Processing*, Institute of Electrical and Electronics Engineers, New York, 1979.
9. J. S. Bendat and A. G. Piersol, *Random Data, Analysis and Measurement Procedures*, Wiley, New York, 1986.
10. B. A. Blesser, "Digitization of Audio: A Comprehensive Examination of Theory, Implementation, and Current Practice," *J. Audio Eng. Soc.*, Vol. 26, 1978, pp. 739–771.
11. C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, Vol. 37, 1949, pp. 10–31.
12. J. C. Burgess, "Digital Spectrum Analysis of Periodic Signals," *J. Acoust. Soc. Am.*, Vol. 58, 1975, pp. 556–567.
13. J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing*, Macmillan, New York, 1988.

14. A. P. G. Peterson, *Handbook of Noise Measurement*, 9th ed., GenRad, Concord, MA, 1980.
15. E. B. Magrab and D. S. Blomquist, *The Measurement of Time-Varying Phenomena*, Wiley, New York, 1971.
16. J. Meyer, "Time Correction of Anti-Aliasing Filters in Digital Audio Systems," *J. Audio Eng. Soc.* (Engineering Reports), 1984, pp. 132–137.
17. R. Kuc, *Introduction to Digital Signal Processing*, McGraw-Hill, New York, 1988.
18. T. F. Darling and M. O. J. Hawksford, "Oversampled Analog-to-Digital Conversion for Digital Audio Systems," *J. Audio Eng. Soc.*, Vol. 38, 1990, pp. 924–943.
19. M. W. Hauser, "Principles of Oversampling A/D Conversion," *J. Audio Eng. Soc.*, Vol. 39, 1991, pp. 3–26.
20. F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, Vol. 66, 1978, pp. 51–83.
21. J. C. Burgess, "Optimum Approximations to Dolph-Chebyshev Data Windows," *IEEE Trans. Signal Proc.*, Vol. 40, 1992, pp. 2592–2594.
22. A. H. Nuttall, "Some Windows with Very Good Sidelobe Behavior," *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-29, 1981, pp. 84–91.
23. J. C. Burgess, "Accurate Analysis of Periodic Signals Using a DFT," in preparation.
24. J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, Vol. 63, 1975, pp. 561–584; reprinted with corrections in D. G. Childers (Ed.), *Modern Spectral Analysis*, Institute of Electrical and Electronics Engineers, New York, 1978.
25. L. J. Eriksson, M. C. Allie, C. D. Bremigan, and J. A. Gilbert, "Active Noise Control on Systems with Time-Varying Sources and Parameters," *Sound Vib.*, July 1989, pp. 16–21.
26. S. Braun and J. K. Hammond, "Additional Techniques," in S. Braun (Ed.), *Mechanical Signature Analysis*, Academic, Orlando, FL, 1986, Chapter 6.
27. C. L. Nikias and J. M. Mendel (Eds.), "Special Section on Higher Order Spectral Analysis," *IEEE Trans. Acoust. Speech Signal Proc.*, Vol. ASSP-38, 1990, pp. 1236–1317.
28. O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *IEEE Signal Proc. Mag.*, Vol. 8, 1991, pp. 14–38.
29. F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and Quadratic Time-Frequency Signal Representatives," *IEEE Signal Proc. Mag.*, Vol. 9, 1992, pp. 21–67.
30. W. A. Gardner, "Exploitation of Spectral Redundancy in Cyclostationary Signals," *IEEE Signal Proc. Mag.*, Vol. 8, 1991, pp. 14–36.
31. M. R. Spiegel, *Theory and Problems of Statistics*, Schaum's Outline Series, McGraw-Hill, New York, 1961.
32. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, Dover, New York, 1958.
33. R. C. Weast and S. M. Selby (Eds.), *Handbook of Tables for Mathematics*, 4th rev. ed., CRC Press, Cleveland, OH, 1975.
34. A. H. Nuttall and G. C. Carter, "Spectral Estimation Using Combined Time and Lag Weighting," *Proc. IEEE*, Vol. 70, 1982, pp. 1115–1125.
35. E. A. Sloane, "Comparison of Linearly and Quadratically Modified Spectral Estimates of Gaussian Signals," *IEEE Trans. Audio Electroacoust.*, Vol. 17, 1969, pp. 133–137.

102

ACOUSTIC HOLOGRAPHY

JULIAN D. MAYNARD

1 INTRODUCTION

Acoustic holography[1] refers to the recording of sound waves in a two-dimensional pattern (the hologram) and the use of the record to reconstruct the entire sound field throughout a three-dimensional region of space; portions of the region that show large amplitudes provide a three-dimensional image of the sources of the sound. The technique is one of the most powerful within the general area of acoustical imaging; its great power arises from the enormous increase in the amount of observable information that occurs when one converts the two-dimensional hologram into the three-dimensional reconstruction. For example, a digitized acoustic hologram may contain a few thousand data points, but the three-dimensional reconstruction may contain many millions of data points. The fundamental physical and mathematical principles involved in holography include the knowledge that the field being measured is linear and obeys the differential equation for wave propagation and that data on a two-dimensional surface are sufficient to determine uniquely the solution to the wave equation. This chapter presents the detailed theory of holography, with which subsequent versions may be understood. Some features of holographic measurement systems will also be presented.

2 DESCRIPTION AND BRIEF HISTORY

Important applications[1] of acoustic holography may be found wherever there is a need to visualize objects or fields that cannot be observed optically, such as objects in an opaque fluid or sound waves in air. Applications include medical imaging, nondestructive evaluation, underwater imaging and survey, seismic surveying, and noise measurement and control.

Acoustic holography is analogous to optical holography, in which light waves are recorded on a two-dimensional film and three-dimensional visual images are reconstructed. However, acoustic holography is free from some technical limitations that constrain optical holography. For example, conventional optical holography is usually limited to recording a single coherent wavelength (or single frequency) of light, whereas acoustic holography can record a wideband (multifrequency) noise field. Furthermore, conventional holography can only reconstruct images involving the amplitude of the light field with restricted resolution (ability to distinguish fine detail on a scale relative to the wavelength of the radiation), but acoustic holography can produce high-resolution images[2] of not only the sound pressure field but also of the motion of the air and of the surface of the source, as well as the flow of acoustic energy around and away from the source.

The reason that acoustic holography is able to surpass conventional optical holography lies in the different methods of recording and processing the hologram data. In conventional optical holography the extremely rapid oscillations of the light waves can be recorded only through the use of a reference beam of light and the phenomenon of optical interference, whereas in acoustic holography microphones readily record both amplitude and phase (with all frequencies and the phase information actually contained in the real-time recording). In optical holography the reconstruction of the three-dimensional images is accomplished in an approximate manner through the physical phenomenon of optical diffrac-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

tion, whereas in acoustic holography the data can be digitized so that it can be processed with a computer with no intrinsic limitations in frequency or resolution. Note that the real-time recording of the acoustic field contains all frequencies and phase information without using a reference beam of sound; that a reference beam is necessary to reconstruct a wave field is a misconception carried over from conventional optical holography.

Many of the early techniques[1] for acoustic holography used reference beams, interference on the surface of water, or other interferometric schemes. Some methods converted the acoustical information into light (with arrays of light sources), recorded the patterns with film, and then reconstructed the holograms with lasers as in optical holography. Most contemporary methods of acoustic holography record microphone signals digitally and perform reconstructions numerically with high-speed computers. The experimental methods are relatively straightforward; one simply tries to record time-series data at a high density of points in space over an aperture that subtends as large a solid angle from the source as possible. Measurements may be made in reverberant as well as anechoic spaces. Because of the use of numerical reconstructions, the mathematical foundations are of crucial importance.

3 MATHEMATICAL FOUNDATIONS FOR HOLOGRAPHY

In mathematical terms, the problem of holography is to find the field $\psi(\mathbf{r}, t)$, a function of time t in a three-dimensional region of space $\{\mathbf{r}\}$, from the field measured on a two-dimensional surface $\psi(\mathbf{r}_H, t)$, given that away from the sources the field obeys the homogeneous wave equation

$$\nabla^2\psi - \frac{1}{c}\frac{\partial^2\psi}{\partial t^2} = 0, \tag{1}$$

where c is the sound speed. In particular, one wishes to find the field $\psi(\mathbf{r}_S, t)$, and other quantities calculated from ψ, evaluated at the surface of a source $\{\mathbf{r}_S\}$, which is usually a vibrating surface, with a surface velocity field $v_n(\mathbf{r}_S, t)$.

The first step is to transform from the time domain (t) to the frequency domain (ω):

$$\psi(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} \psi(\mathbf{r}, t)e^{i\omega t}\, dt, \tag{2}$$

$$\psi(\mathbf{r}_H, \omega) = \int_{-\infty}^{\infty} \psi(\mathbf{r}_H, t)e^{i\omega t}\, dt. \tag{3}$$

The wave equation becomes the Helmholz equation[3]

$$\nabla^2\psi(\mathbf{r}, \omega) + k^2\psi(\mathbf{r}, \omega) = 0, \tag{4}$$

where $k = \omega/c$. The surface velocity field $v_n(\mathbf{r}_S, \omega)$ can now be related to the normal derivative of the field ψ with

$$\frac{\partial\psi}{\partial n} \equiv \nabla\psi(\mathbf{r}_S, \omega)\cdot\hat{n} = i\omega\rho v_n(\mathbf{r}_S, \omega), \tag{5}$$

where ρ is the mass density of the fluid and $\hat{n}$ is a unit vector normal to the surface at $\mathbf{r}_S$; $\partial\psi/\partial n$ is a simplified notation for the normal derivative.

It should be noted that formally it is necessary to have fields evaluated continuously for all times, $-\infty < t < \infty$. In practice this cannot be done; the wave field is sampled at discrete points in time over a finite window in time. How this is handled is covered in standard texts on signal processing.[4] In acoustic holography, the main issue is the processing of the spatial information.

In order to develop the theory for the spatial processing, the problem is first inverted. That is, we assume that the surface velocity (i.e., the normal derivative of ψ at $\mathbf{r}_S$) is known, and then calculate $\psi(\mathbf{r})$ and $\psi(\mathbf{r}_H)$. If we have a field $G(\mathbf{r})$ which satisfies

$$\nabla^2 G(\mathbf{r}) + k^2 G(\mathbf{r}) = -4\pi\delta(\mathbf{r}), \tag{6}$$

where δ is the delta distribution, then Green's theorem yields

$$\psi(\mathbf{r}) = \frac{1}{4\pi}\iint \cdot\left(\psi(\mathbf{r}_S)\frac{\partial G}{\partial n}(\mathbf{r}-\mathbf{r}_S) - i\omega\rho v_n(\mathbf{r}_S)G(\mathbf{r}-\mathbf{r}_S)\right)\cdot d^2 r_S. \tag{7}$$

The problem is that while $v_n(\mathbf{r}_S)$ is known, ψ itself is not known on the surface (the situation is reversed for Dirichlet boundary conditions). There are two ways in which this problem may be solved:

1. One finds a function G such that $\partial G/\partial n = 0$; this is the Green's function for the problem (not the free-space Green's function to be encountered later).[3] This is possible only if the surface $\mathbf{r}_S$ and the holo-

gram surface $\mathbf{r}_H$ are level surfaces of a separable coordinate system. There are 11 such coordinate systems for the Helmholz equation, including Cartesian, cylindrical, spherical, and so forth. This solution is referred to as the *eigenfunction expansion.*

2. One uses the free-space Green's function[3] in Eq. (7), and evaluates the equation for $\mathbf{r} = \mathbf{r}_S$, yielding an integral equation for $\psi(\mathbf{r}_S)$ in terms of $v_n(\mathbf{r}_S)$. This equation, the surface Helmholz integral equation, must be solved numerically. Once $\psi(\mathbf{r}_S)$ is known, then Eq. (7) may be evaluated for any $\mathbf{r}$ (also numerically). This solution is referred to as the *numerical solution.*

If the surface of the actual source is not a level surface of a separable coordinate system, it may be sufficient to replace the source with an equivalent one formed with a level surface that approximates the actual source. In this case one may use the eigenfunction expansion, which has the advantage of being faster to execute than the numerical solution (for a given accuracy).

In the sections below, the two solutions will be described separately.

3.1 The Eigenfunction Expansion

If $\{\mathbf{r}_S\}$ and $\{\mathbf{r}_H\}$ are level surfaces of a separable coordinate system, then it is possible to find the Green's function G (with $\partial G/\partial n = 0$) as an expansion in terms of eigenfunctions.[3] One lets ξ_1, ξ_2, ξ_3 indicate the coordinates of the separable system, with the surface of the source given by the level surface $\xi_3 = \xi_{3S}$. In these terms, the Green's function may be found as discussed in Ref. 3. Then the general solution may be found with three steps[5]: the hologram data set is transformed, multiplied by a kernel, then inverse transformed. While the solution can be expressed formally in terms of general eigenfunctions of separable curvilinear coordinate systems, it is easiest to treat each of the coordinate systems separately, using properties of special functions available in textbooks. The treatment of cartesian, cylindrical, and spherical coordinates are presented in the next sections.

Cartesian Coordinates In Cartesian coordinates $(\xi_1, \xi_2, \xi_3) = (x, y, z)$, and the level surfaces are planes. For the hologram plane we take $z = z_H$, and the source plane is $z = z_S$, with $z_S < z_H$. The eigenfunctions are

$$U(x, y, z; k_x, k_y) = e^{i(k_x x + k_y y)} e^{i\tilde{k}_z z}, \tag{8}$$

where

$$\tilde{k}_z = \begin{cases} (k^2 - k_x^2 - k_y^2)^{1/2} & \text{if } k_x^2 + k_y^2 < k^2, \\ i(k_x^2 + k_y^2 - k^2)^{1/2} & \text{otherwise.} \end{cases} \tag{9}$$

The transformation to the two-argument function space (indicated with a caret) is

$$\hat{\psi}(k_x, k_y; z_H) = \iint \psi(x, y, z_H) e^{-i(k_x x + k_y y)}\, dx\, dy, \tag{10}$$

which is the ordinary two-dimensional spatial Fourier transform. The solution becomes

$$\psi(x, y, z) = \iint \hat{\psi}(k_x, k_y; z_H) e^{i\tilde{k}_z(z - z_H)} e^{i(k_x x + k_y y)} \cdot dk_x\, dk_y. \tag{11}$$

Letting z go to $z_S < z_H$ reconstructs the source plane, and $z > z_H$ can be used to calculate the far-field properties. It should be noted that the quantity $\tilde{k}_z$ must be handled carefully in order to properly account for both propagating and evanescent wave eigenfunctions.[2]

Equations (10) and (11) involve integrals of continuous functions over infinite domains, which cannot be evaluated in practice. Thus the domains must be made discrete and finite with some procedure. The simplest procedure is to assume that the field is piecewise constant over patches of size $(L/N) \times (L/N)$ over a spatial window of size $L \times L$, and is zero outside the window. The patches are labeled with pairs of integers l, $m = 0, 1, \ldots, N-1$. Equation (10) is replaced with a discrete Fourier transform (DFT):

$$\hat{\psi}_{\mu\nu}(z_H) = \sum_{l=0}^{2N-1} \sum_{m=0}^{2N-1} \psi(x_l, y_m, z_H) e^{-i\pi(l\mu + m\nu)/N}, \tag{12}$$

where the upper limits are extended from $N - 1$ to $2N - 1$ to include a guard band of zeroes.[6] Equation (11) becomes

$$\psi(x_l, y_m, z) = \sum_{\mu=0}^{2N-1} \sum_{\nu=0}^{2N-1} \hat{\psi}_{\mu\nu}(z_H) g_{\mu\nu}(z - z_H) \cdot e^{i\pi(l\mu + m\nu)/N}, \tag{13}$$

where $g_{\mu\nu}(z) = \exp(i\tilde{k}_z z)$, with values for k_x (in $\tilde{k}_z$) given by

$$k_x = \begin{cases} \pi\mu/L & \text{if } 0 \le \mu < N, \\ -\pi(2N - \mu)/L & \text{if } N \le \mu < 2N, \end{cases} \tag{14}$$

and a similar expression for k_y. There are other possible expressions for $g_{\mu\nu}$ that reduce the errors caused by the approximation of the continuous, infinite integrals with discrete, finite functions. Another source of error arises from the possibility that the evanescent wave modes may decay to values below the "noise" floor of the hologram; when these modes are reconstructed, they are multiplied by an exponentially large inverse propagation factor, which may greatly amplify the noise. This problem may be alleviated by multiplying the higher spatial frequency evanescent wave modes with a windowing function. A complete discussion of the details involved in holography for cartesian coordinates, including errors, different kernels, "zoom imaging," and so forth, is presented in the study by Veronesi and Maynard.[6]

Cylindrical Coordinates In cylindrical coordinates $(\xi_1, \xi_2, \xi_3) = (r, \phi, z)$, where r is the perpendicular distance from the z-axis, and ϕ is the azimuthal angle. The hologram is measured on the cylindrical surface $r = r_H$. The eigenfunctions are

$$U(r, \phi, z; m, k_z) = H_m(\tilde{k}_r r) e^{im\phi} e^{ik_z z}, \tag{15}$$

where m is an integer, $\tilde{k}_r = \sqrt{k^2 - k_z^2}$, $H_m(\tilde{k}_r r)$ is the Hankel function when the argument is real, and the modified Hankel function when the argument is imaginary (corresponding to propagating and evanescent waves, respectively). The transformation to the two-argument function space is

$$\hat{\psi}(m, k_z; r_H) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dz \int_0^{2\pi} d\phi\, \psi(r_H, \phi, z) \cdot e^{-im\phi} e^{-ik_z z}. \tag{16}$$

The solution becomes

$$\psi(r, \phi, z) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{\psi}(m, k_z; r_H) \cdot e^{im\phi} e^{ik_z z} \frac{H_m(\tilde{k}_r r)}{H_m(\tilde{k}_r r_H)}\, dk_z. \tag{17}$$

As for the Cartesian coordinate system, the integrals and sums for the cylindrical coordinate system may be made discrete and finite.

Spherical Coordinates In spherical coordinates $(\xi_1, \xi_2, \xi_3) = (r, \theta, \phi)$ where r is the distance from the origin, θ is the polar angle, and ϕ is the azimuthal angle. The hologram is measured on the spherical surface $r = r_H$. The eigenfunctions are

$$U(r, \theta, \phi; l, m) = Y_{lm}(\theta, \phi) h_l(kr), \tag{18}$$

where l and m are integers (with $-l \le m \le l$), $Y_{lm}(\theta, \phi)$ are the spherical harmonics, and $h_l(kr)$ are the spherical Bessel functions. The transformation to the two-argument function space is

$$\hat{\psi}(l, m; r_H) = \int_0^{\pi} \sin\theta\, d\theta \int_0^{2\pi} d\phi\, \psi(r_H, \theta, \phi) Y_{lm}^*(\theta, \phi), \tag{19}$$

where the asterisk indicates the complex conjugate. The solution becomes

$$\psi(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{\psi}(l, m; r_H) Y_{lm}(\theta, \phi) \frac{h_l(kr)}{h_l(kr_H)}. \tag{20}$$

Eigenfunctions for Other Situations In the preceding sections it was assumed that the holography was performed in free space or under anechoic conditions. It is also possible to use acoustic holography in reverberant spaces, using either of two methods. One method[7] is to make the holography measurements on two surfaces (usually closely spaced), and with the two data sets one can separate the outgoing and incoming wave fields. The second method[8] is to make the usual hologram measurement on one surface, but analyze the data using eigenfunctions appropriate for the reverberant space.

Calculation of Other Quantities Once the solution $\psi(\xi_1, \xi_2, \xi_3)$ has been found for the three-dimensional domain, expressions for all other quantities may be found.[9] In particular, expressions for the components of the particle velocity in the coordinate directions may be found using an equation of the form

$$v_j(\xi_1, \xi_2, \xi_3) = -\frac{1}{i\omega\rho} \frac{\partial\psi}{\partial\xi_j}, \tag{21}$$

and the derivatives may be carried out explicitly inside the integrals or sums in Eqs. (11), (17), and (20). The acoustic vector intensity may be calculated using

$$S_j(\xi_1, \xi_2, \xi_3) = \tfrac{1}{2}\mathrm{Re}\{\psi(\xi_1, \xi_2, \xi_3) v_j^*(\xi_1, \xi_2, \xi_3)\}. \tag{22}$$

An expression for the total power radiated may be found by integrating Eq. (22) explicitly using properties of the eigenfunctions. The expression for the total power for the Cartesian coordinate system is

$$P = \frac{1}{8\pi^2 \rho c} \iint_{k_x^2 + k_y^2 \le k^2} |\hat{\psi}(k_x, k_y; z_H)|^2 \cdot (1 - (k_x^2 + k_y^2)/k^2)^{1/2} \, dk_x \, dk_y. \tag{23}$$

It is interesting to note that in the expressions for the total power, the evanescent wave components are eliminated, as they must be.

The far-field directivity pattern $D(\theta, \phi)$ may also be calculated. For the Cartesian system this is given by

$$D(\theta, \phi) = k \cos\theta \, \hat{\psi}(k \sin\theta \cos\phi, k \cos\theta \cos\phi, z_H). \tag{24}$$

3.2 Numerical Solution for Arbitrarily Shaped Sources

For sources whose surfaces may not be represented by a level surface of a separable coordinate system, the Green's function must be determined numerically. A procedure for accomplishing this was developed by Veronesi and Maynard.[10,11] In Eq. (7), one must use the free-space Green's function, defined by

$$G(\mathbf{r}) = \frac{e^{ikr}}{r}, \tag{25}$$

where r is the length of the vector $\mathbf{r}$, and the unknown term in Eq. (7), $\psi(\mathbf{r}_S)\partial G/\partial n$, must be determined numerically. This is done by evaluating Eq. (7) for $\mathbf{r}$ on the surface $\mathbf{r}_S$. Because of the singular nature of G, a factor of $\frac{1}{2}$ appears when a principal value is taken (for sharp edges or corners, the factor is different[11]):

$$\tfrac{1}{2}\psi(\mathbf{r}_S) = \frac{1}{4\pi} \iint \psi(\mathbf{r}') \frac{\partial G}{\partial n} (\mathbf{r}_S - \mathbf{r}') \, d^2 r' - \frac{i\omega\rho}{4\pi} \iint v_n(\mathbf{r}') G(\mathbf{r}_S - \mathbf{r}') \, d^2 r', \tag{26}$$

which is the surface Helmholtz integral equation. The integrations in this equation must be replaced by appropriate numerical integration approximations to solve the general problem of odd-shaped radiators. At some point in their application, most numerical integration schemes reduce to matrix operations on a set of coefficients representing the fields $v_n(\mathbf{r}_s)$ and $\psi(\mathbf{r}_s)$. Denoting these coefficient sets with indices such as in $v_n(\mathbf{r}_s)_i$ or $\psi(\mathbf{r}_s)_q$, the matrix approximation to Eq. (26) is

$$\tfrac{1}{2}\psi(\mathbf{r}_S)_q = D_{qj}\psi(\mathbf{r}_S)_j - M_{qj} v_n(\mathbf{r}_S)_j \tag{27}$$

with summation implied over all repeated indices. Equation (27) can be rewritten as

$$\overline{D}_{qj}\psi(\mathbf{r}_S)_j = M_{qj} v_n(\mathbf{r}_S)_j, \tag{28}$$

where $\overline{D}_{qj} = D_{qj} - \frac{1}{2} I_{qj}$, $[I]$ is the identity matrix, and the matrices $[D]$ and $[M]$ are appropriate to the chosen numerical integration scheme. A typical scheme is presented by Koopman and Benner.[12]

At certain characteristic frequencies, Eq. (28) is not sufficient to determine a unique solution to the radiation problem.[12–14] The problem is that, at these frequencies, interior eigenmodes exist that are homogeneous solutions over the surface under study. To any particular solution of the problem, it is then possible to add one or more homogeneous solutions. This problem manifests itself as nearly singular matrices $[D]$ and $[M]$ in the approximation of Eq. (26). The homogeneous solutions of the surface Helmholtz integral equation are not radiating solutions, and it is possible to augment the surface equation with an interior equation to eliminate these solutions.[13] Written in terms of the surface acoustic field $\psi(\mathbf{r}_s)$ and the surface velocity $v_n(\mathbf{r}_S)$, the interior equation is

$$0 = \frac{1}{4\pi} \iint \psi(\mathbf{r}') \frac{\partial G}{\partial n} (\mathbf{r}_I - \mathbf{r}') \, d^2 r' - \frac{i\omega\rho}{4\pi} \iint v_n(\mathbf{r}') G(\mathbf{r}_I - \mathbf{r}') \, d^2 r', \tag{29}$$

where the point $\mathbf{r}_I$ is interior to the surface. Equation (29) is the interior Helmholz integral equation. It can be approximated by

$$D_{qj}\psi(\mathbf{r}_S)_j = M_{qj} v_n(\mathbf{r}_S)_j, \tag{30}$$

where the subscript q's now refer to a number of points $\mathbf{r}_I$ in the interior. Equations (28) and (30) can be treated as an overdetermined system of equations and solved accordingly.

The problem of determining the characteristic frequencies of an odd-shaped surface is equivalent to measuring the relative singularity of the matrices as a function of frequency. According to Stewart,[15] the difficulty in determining numerical singularities is that the matrices involved are only approximations to the actual problem under study and so may only be close to singular matrices when the real problem is singular. Furthermore, many methods for decomposing a matrix may transform a nearly singular matrix to a form that is not nearly singular. However, the singular-value decomposition discussed later yields a nearly singular matrix if and only if the original matrix is nearly singular.[15] Therefore, this decomposition is ideally suited for determining the characteristic frequencies of an odd-shaped surface and the eigenmodes as well. This technique was originally introduced by Veronesi and Maynard.[10]

By premultiplying and postmultiplying a general matrix by a set of unitary transformations, the singular-value decomposition reduces the matrix to a diagonal form. By gathering these transformations into matrices $[U]$ and $[V^\dagger]$ (the dagger indicates the conjugate transpose), the general $m \times n$ matrix $[X]$, where $m > n$, can be written as

$$[X] = [U]\begin{pmatrix} \Sigma \\ 0 \end{pmatrix}[V^\dagger], \tag{31}$$

where $[U][U^\dagger] = [I], [V][V^\dagger] = [I]$, and $[\Sigma]$ is an $n \times n$ diagonal matrix with diagonal elements δ_k, the so-called singular values.[15] It can be readily shown that the singular values measure the relative output vector magnitude for each of a complete set of input vectors, namely the columns of $[V]$. (For a hermitian matrix, the singular values are the eigenvalues, and the columns of $[V]$ are the eigenvectors). In decomposing either $[M]$ or $[D]$, this set covers all possible surface velocity or pressure distribution approximations. It is the relatively low-frequency, long-wavelength regime that is often of interest in acoustic holography, and it is assumed that, if interior eigenmodes exist, they are few in number relative to the number of degrees of freedom of the approximation scheme. Since all of the columns of $[V]$ are orthogonal, eigenmodes, if they exist, must be represented by a distinct set of columns in $[V]$ possessing correspondingly small singular values in comparison with the singular values of the majority of other modes. Therefore, the presence of a very small singular value, relative to the remainder of values for a given matrix, indicates the existence of an interior eigenmode, at least in a numerical sense. In order to determine the interior eigenfrequencies, only the singular values need be calculated, and this can be done in much less time then is required for the complete decomposition as in the form of Eq. (31).

The goal now is to form an invertible matrix that relates the surface pressure and normal component of surface velocity as follows:

$$\psi(\mathbf{r}_S)_q = S_{qj} v_n(\mathbf{r}_S)_j \,. \tag{32}$$

At noncharacteristic frequencies, the matrices of Eq. (28) should be nonsingular, and the matrix $[S]$ may be formed directly as

$$S_{qj} = \overline{D}_{qn}^{-1} M_n. \tag{33}$$

At characteristic frequencies, the matrix must be generated from the overdetermined system of Eqs. (28) and (30). The overdetermined system may be written in the form of a single matrix equation as

$$\begin{pmatrix} \overline{D} \\ -D' \end{pmatrix}[p(\mathbf{r}_s)] = \begin{pmatrix} M \\ M' \end{pmatrix}[v_n(\mathbf{r}_s)]. \tag{34}$$

An equation with a single $n \times n$ matrix, where n is the number of degrees of freedom of the surface model [as in Eq. (32)], may be obtained from Eq. (34) by utilizing the singular-value decomposition to form a generalized inverse of the left-hand side matrix in Eq. (34).[15] The decomposition has the form

$$\begin{pmatrix} \overline{D} \\ -D' \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}\begin{pmatrix} \Sigma \\ 0 \end{pmatrix}[V^\dagger]. \tag{35}$$

The generalized inverse of this matrix is given by

$$\begin{pmatrix} \overline{D} \\ -D' \end{pmatrix}^{-1} = [V][\Sigma^{-1}0]\begin{pmatrix} U_{11}^\dagger & U_{12}^\dagger \\ U_{21}^\dagger & U_{22}^\dagger \end{pmatrix}. \tag{36}$$

This equation reduces to the form

$$\begin{pmatrix} \overline{D} \\ -D' \end{pmatrix}^{-1} = [V][\Sigma^{-1}][U_{11}^\dagger U_{12}^\dagger]. \tag{37}$$

A similar decomposition and reduction can be performed on the right-hand side matrix in Eq. (34) yielding the following equation:

$$\begin{pmatrix} M \\ M' \end{pmatrix} = \begin{pmatrix} W_{11} \\ W_{12} \end{pmatrix}[\Gamma][Z^\dagger], \tag{38}$$

where $[\Gamma]$ is a diagonal matrix and $[W_{ij}]$ and $[Z]$ are unitary matrices. The matrix $[S]$ is then given by the expression

$$[S] = [V][\Sigma^{-1}][U_{11}^{\dagger} \quad U_{12}^{\dagger}]\begin{pmatrix} W_{11} \\ W_{12} \end{pmatrix}[\Gamma][Z^{\dagger}]. \qquad (39)$$

When $[S]$ is used in Eq. (32), the surface problem is solved.

In order to reduce the size of the numerical odd-shaped surface reconstruction problem, it is desirable to form a single matrix that directly relates the normal component of surface velocity to the pressure at a number of points in the field above the surface. This matrix is referred to as the field matrix $[F]$, and with it the following equation can be written:

$$\psi(\mathbf{r})_q = F_{qj}v_n(\mathbf{r}_S)_j\,. \qquad (40)$$

For odd-shaped sources for which the Green's function is not known, there is no single integral relating the surface velocity and field pressure, although the pressure is uniquely determined by Eqs. (7), (26), and (29). The method of reduction of Eqs. (26) and (29) to finite form can also be applied to the reduction of Eq. (7) to a finite and discrete form:

$$\psi(\mathbf{r}_q) = \tfrac{1}{2}D_{qj}\psi(\mathbf{r}_S)_j - \tfrac{1}{2}M_{qj}v_n(\mathbf{r}_S)_j\,, \qquad (41)$$

where the matrices $[D]$ and $[M]$ are defined as in Eqs. (32) and (33) with $\mathbf{r}_q$ being a field point instead of a surface point.

By using the surface solution matrix $[S]$ of Eq. (39), the surface pressure vector $\psi(\mathbf{r}_s)$ can be written in terms of the surface velocity. With this replacement, Eq. (41) can be written as

$$\psi(\mathbf{r})_q = \tfrac{1}{2}(D_{qn}S_{nj} - M_{qj})v_n(\mathbf{r}_S)_j\,, \qquad (42)$$

and the elements of $[F]$ can be identified as

$$F_{qj} = \tfrac{1}{2}(D_{qn}S_{nj} - M_{qj}). \qquad (43)$$

Equation (43) provides the formula for constructing the elements of $[F]$ from a set of readily calculated coefficients for any odd-shaped geometry.

Given the normal component of surface velocity for the odd-shaped source and the surface and interior Helmholtz integral equations, a unique surface pressure field is determined. These two surface fields in turn define a unique pressure field above the surface by way of Eq. (7). However, this radiated pressure field can only be observed experimentally over a limited region, with finite resolution, and with finite dynamic range. In analogy with the theory of holography in Cartesian coordinates, it is intuitive that high spatial frequency modes of an odd-shaped surface must generate evanescent fields when their wavenumber exceeds that of the propagating wavenumber. The information about these modes is diminished at the field measurement loci relative to the information from lower spatial frequency radiating modes. If the evanescent field of one mode falls below the dynamic range of the measuring system relative to a second mode, surface modes with or without this evanescent mode will appear experimentally identical. Numerically this problem manifests itself as Eq. (42) being singular. The singular-value decomposition technique is again indispensable for inverting this matrix equation.[10,11] The procedure is as follows:

The singular values of the matrix $[F]$ are found, and relatively small singular values are identified as representing pathological, weakly radiating modes, which may fall below the "noise" floor of the hologram. The small singular values are replaced by the largest number available in the computer, and the matrix $[F]$ is inverted. The holographic reconstruction matrix $[F]^{-1}$ is now biased against the pathological modes.

4 ACOUSTIC HOLOGRAPHIC MEASUREMENT SYSTEMS

As mentioned in the introduction, contemporary acoustic holography involves mostly data processing, with the experimental measurement being relatively straightforward. Typical contemporary acoustic holographic measurement systems may be found in the references. One example is presented below.

A schematic representation of an acoustic holographic system[9] is illustrated in Fig. 1. The recording "medium" consists of a large array of microphones. The array is constructed with a square frame containing a 16 × 16 latticework of wires; the microphones are located at the intersections of the wires, resulting in an array of 256 microphones. The sound source to be studied is positioned just below the microphone array. The microphone signals are routed through a high-speed multiplexer and digitizer to a computer memory, which constitutes, in this case, the recorded hologram. The microphones are less than 0.01 m in diameter, and the spacing between microphones is 0.18 m, so that the array is very transparent. The array is nearly 3 m^2, and since it is positioned close to the source plane, it subtends a large solid angle rela-

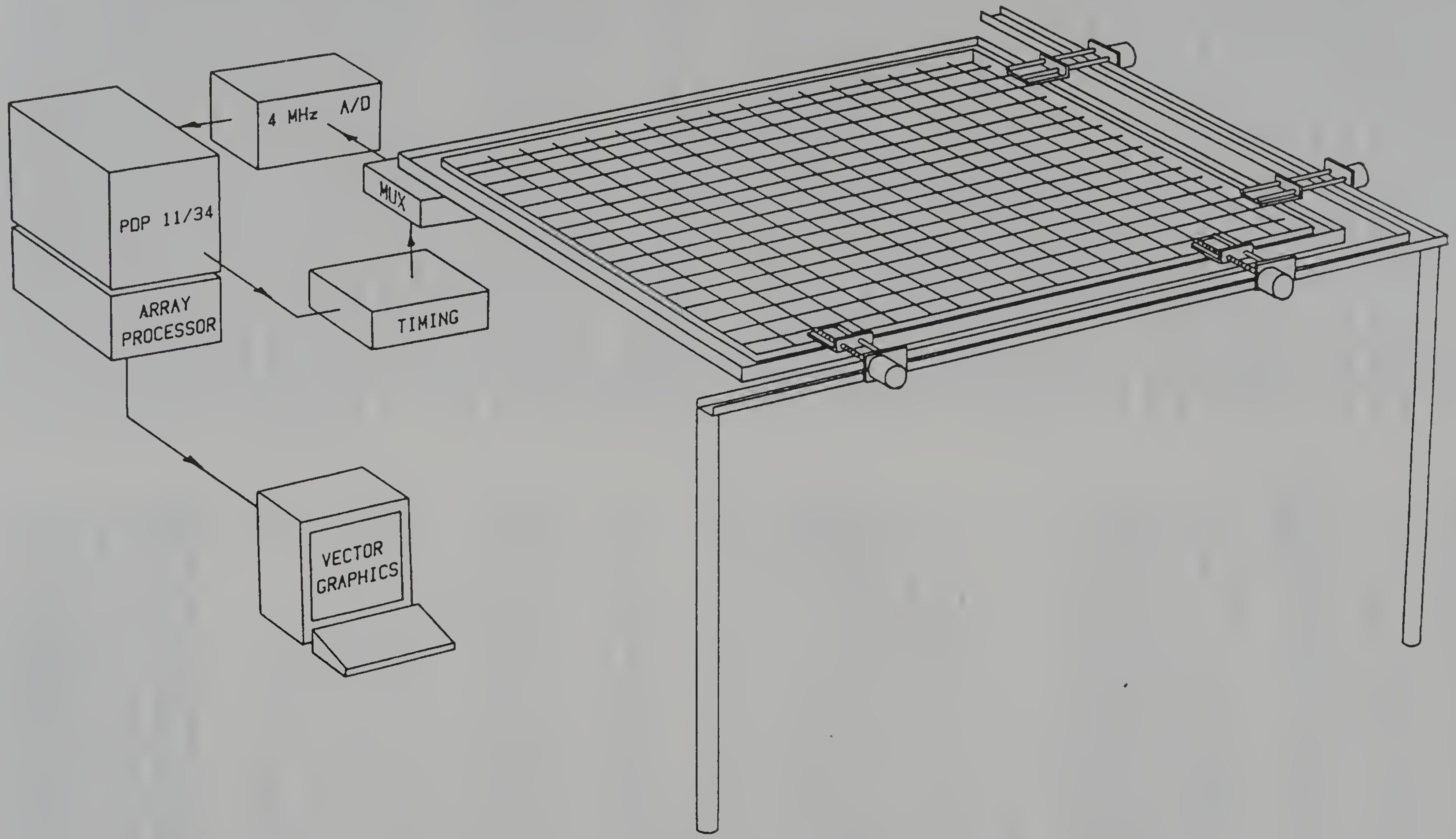

PDP 11/34
ARRAY PROCESSOR
4 MHz A/D
MUX
TIMING
VECTOR GRAPHICS

Fig. 1 Schematic illustration of an acoustical holographic system based on a 16 × 16 array of microphones. The sound source is positioned below the microphone array, and the microphone signals are processed with a computer.

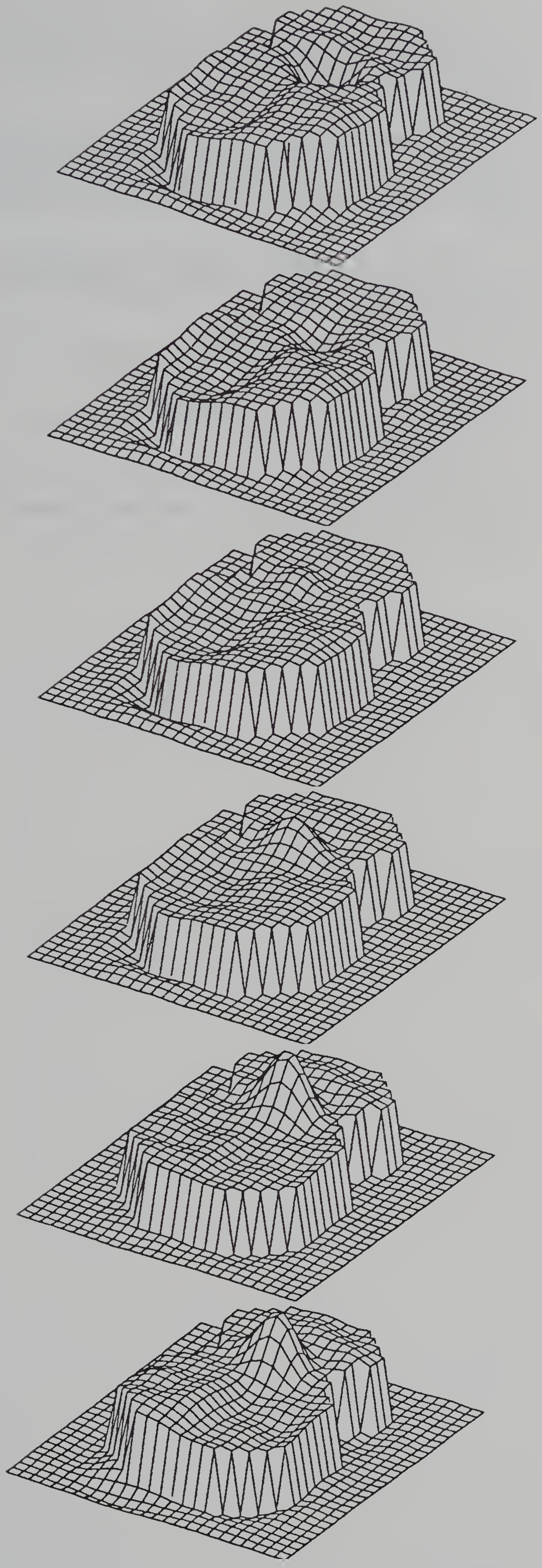


Fig. 2 Example of a sequence of acoustic holographic reconstructions showing the motion (exaggerated) of the soundboard of the guitar as well as the motion of the air moving in and out of the sound hole.

tive to the sources. Further details of the system may be found in Ref. 9.

The recorded hologram is processed by the computer in the following manner: The data are Fourier transformed in time to separate the different temporal frequency components; each frequency component is then Fourier transformed spatially; the transformed data are multiplied by the kernel, which contains the information about how the sound waves must travel (according to the wave equation); the results are processed with an inverse Fourier transform; and finally the results are displayed with graphic routines that produce three-dimensional "hidden line" displays, energy flow maps, and so forth.

Because of the freedom afforded through the use of a computer to accomplish acoustic holographic reconstructions, the images produced may be magnified or exaggerated in order to enhance some feature one would like to observe. Furthermore, since the time variation of the sound has been included in the analysis of the data, it is possible to produce motion picture images of the sound radiation and its generation by the source. As an example, Fig. 2 shows a sequence of acoustic holograms reconstructing the surface of a guitar as it vibrates and radiates sound; the frames show, as an evolution in time, the motion (greatly exaggerated) of the soundboard of the guitar as well as the motion of the air moving in and out of the sound hole.

There are other systems that involve the principles of acoustic holography, and these need not be as complete and detailed as the system described above. In particular, there are a number of systems[7,8,16] that use only one or a few microphones, and scan in order to obtain the necessary spatial information. It should be noted that these require that the source be steady or under the control of the experimenter or that a stationary reference microphone be used.

REFERENCES

1. A. F. Metherell et al., *Acoustical Holography*, Vols. 1–7, and *Acoustical Imaging* Vols. 8–14, Plenum, New York, 1969–1985.
2. E. G. Williams and J. D. Maynard, "Holographic Imaging without the Wavelength Resolution Limit," *Phys. Rev. Lett.*, Vol. 45, No. 7, 1980, pp. 554–557.
3. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, 1978, pp. 803–834.
4. A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, New York, 1962.
5. M. M. Sondhi, "Reconstruction of Objects from Their Sound Diffraction Pattern," *J. Acoust. Soc. Am.*, Vol. 46, No. 2, 1969, pp. 1158–1164.
6. W. A. Veronesi and J. D. Maynard, "Nearfield Acoustic

Holography (NAH) II. Holographic Reconstruction Algorithms and Computer Implementation," *J. Acoust. Soc. Am.*, Vol. 81, No. 5, 1987, pp. 1307–1322.

7. G. Weinreich and E. B. Arnold, "Method for Measuring Acoustic Radiation Fields," *J. Acoust. Soc. Am.*, Vol. 68, No. 2, 1980, pp. 404–411.
8. M. Villot, G. Chaveriat, and J. Roland, "Phonoscopy: An Acoustical Holography Technique for Plane Structures in Enclosed Spaces," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 187–195.
9. J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield Acoustic Holography (NAH) I. Theory of Generalized Holography and the Development of NAH," *J. Acoust. Soc. Am.*, Vol. 78, No. 4, 1985, pp. 1395–1413.
10. W. A. Veronesi and J. D. Maynard, "Acoustic Holography for Odd-Shaped Sources," in R. Lotz (Ed.), *Proceedings Inter-Noise 86*, Noise Control Foundation, New York, 1986, pp. 1251–1256.
11. W. A. Veronesi and J. D. Maynard, "Digital Holographic Reconstruction of Sources with Arbitrarily Shaped Surfaces," *J. Acoust. Soc. Am.*, Vol. 85, No. 2, 1989, pp. 588–598.
12. G. H. Koopmann and H. Benner, "Method of Computing the Sound Power of Machines Based on the Helmholtz Integral," *J. Acoust. Soc. Am.*, Vol. 71, No. 1, 1982, pp. 78–89.
13. H. A. Schenck, "Improved Integral Formulation for Acoustic Radiation Problems," *J. Acoust. Soc. Am.*, Vol. 44, 1967, pp. 41–58.
14. F. Ursell, "On the Exterior Problems of Acoustics," *Proc. Camb. Phil. Soc.*, Vol. 74, 1973, pp. 117–123.
15. G. W. Stewart, *Introduction to Matrix Computations*, Academic, New York, 1973, p. 318.
16. E. G. Williams and H. D. Dardy, "Nearfield Acoustical Holography Using an Underwater Automated Scanner," *J. Acoust. Soc. Am.*, Vol. 78, No. 2, 1985, pp. 789–798.

PART XI

PHYSIOLOGICAL ACOUSTICS

103

INTRODUCTION

JOSEPH E. HIND

1 INTRODUCTION

Physiological acoustics (*PA*) is a subdivision of acoustics involved in the study of structure and function of the auditory system in animals ranging from insects to humans. Although originally defined more narrowly, PA now includes the entire auditory system, that is, all of the acoustical and neural events that occur when a sound stimulus evokes the spatial and temporal patterns of neuronal activity in the brain that underlie sensation and perception in hearing. From a disciplinary standpoint, PA embraces many of the facets of neuroscience, including the subdivision designated neurobiology, which emphasizes basic biological processes, especially cellular and molecular mechanisms. Contemporary research in PA typically involves interdisciplinary teams whose members come from such diverse disciplines as acoustics, anatomy, audiology, biochemistry, biophysics, computer science, electrical engineering, neurophysiology, molecular biology, otology, psychology and pharmacology.

2 ORIENTATION OF CHAPTERS IN PART XI

This and the 10 following chapters in Part XI review knowledge and working hypotheses in PA from several perspectives. "Working hypotheses" are included to emphasize the relatively short half-life of beliefs and ideas in PA in comparison with, for example, the laws of physical acoustics. Several chapters restrict consideration to a specific locus in the auditory pathway (e.g., external ear, middle ear, cochlea) while others are oriented according to discipline (anatomy, biochemistry, pharmacology). Chapters 104 and 113 have been written in a style and with an objective not typically found in a handbook. Whereas each of the other chapters is primarily intended to act as a source of facts for a relatively restricted segment of PA, Chapters 104 and 113 draw upon data from a broad range of observations and, with the aid of some plausible speculation, formulate a mutually consistent set of working hypotheses that explain specific complex functions of auditory systems. Chapter 113 also illustrates the great advantages that can accrue from choosing, as an experimental subject, an animal that manifests such highly specialized behavior and sensory systems as the echo-locating bats.

3 ORGANIZATION OF THIS CHAPTER

This chapter is primarily intended to introduce the field of PA to readers who are not specialists in this area. To provide an indication of the scope of research in PA, the chapter begins by outlining some of the broad categories of *methods* in current use. This is followed by a brief overview of the ways in which the peripheral auditory system responds to sound, tracing the stream of stimulus information from the outer ear, through the middle ear, to the cochlea where the information is coded in the discharges of auditory nerve fibers. *Cochlear function* is one of the most active foci of ongoing research in PA; this topic is addressed in much greater detail in Chapters 107 and 108, but a short summary is included here, both because of its fundamental importance to understanding of the auditory system and as an example of a highly dynamic field in which changes in basic concepts over the past two decades have been nothing less than revolutionary. It is suggested that Chapter 107 be

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

read before consulting any other material on the peripheral auditory system. All readers are encouraged to utilize the detailed information and extensive bibliographic references in each of the remaining 10 chapters in Part XI.

The present chapter next points out some relationships between physiological acoustics and *psychological acoustics*, including the use of *standardized terminology*, which succinctly and unambiguously describes the manner in which auditory stimuli are applied to one or both ears. For purposes of orientation, a short commentary on each of the 10 chapters is then presented. This chapter concludes with a few broadly based references that review and summarize some of the major areas of PA research.

4 METHODS

From the foregoing description of the breadth of PA, it follows that the methods used in present-day research include, in addition to acoustical measurements and theory, virtually the full complement of techniques employed in neuroscience at all levels: molecular, cellular, and systems. A brief listing of some of the more prominent methods follows.

4.1 Acoustics

Acoustical methodology informs and provides crucial support to all areas of research and practical applications in PA. From steadily increasing sophistication in the generation and monitoring of auditory stimuli to characterization of the biological components of the auditory system in terms of systems analysis, very few research strategies in contemporary PA could be pursued in the absence of acoustical instrumentation and theory. As a minor example, a growing number of investigators are studying the merit of measuring the energy absorbed by the ear as an index of auditory stimulus strength rather than specifying only sound pressure as has commonly been done (see Chapters 105 and 106). On a related topic, it may come as a surprise to learn that for high frequencies there is still uncertainty as to the exact modes of vibration of the eardrum and the attached malleus. The malleus is the first in the chain of three ossicles in the mammalian middle ear that conduct the vibratory energy of a stimulus from the eardrum to the fluid of the inner ear (cochlea). Even more challenging problems in vibration measurement arise within the cochlea, ranging from studies of multicellular structures such as the basilar membrane to the stereocilia of an individual hair cell. The most commonly used methods for the measurement of vibration in the ear involve some form of laser interferometry although the Mössbauer technique remains useful in some situations.

A final example of PA research in which acoustical methodology is playing a vital role is the development of virtual acoustic space paradigms. In this technique, earphones are used to generate stimuli at the two eardrums that reproduce the sounds recorded previously at these sites for a free-field sound source at a specified direction with respect to the listener. In areas of applied research the utility of virtual space methods is being investigated in such diverse fields as the entertainment industry and in military efforts to facilitate interaction between an airplane pilot and the surrounding environment. In basic research, virtual space stimuli are proving valuable in studies of directional hearing in human listeners and in electrophysiological studies of the spatial receptive fields of single neurons in experimental animals. Such earphone-generated stimuli are not only much more convenient to use than their free-field counterparts but the method also enables independent control of each of the attributes of the stimuli (e.g., interaural time and intensity differences) in a manner that would not be possible under actual free-field conditions.

4.2 Electrophysiology

Recordings of electrical potentials in both experimental animals and human subjects have occupied a prominent position in PA since the origin of the field. Since a detailed description of electrophysiological studies of the central auditory nervous system is presented in Chapter 111, only a brief listing of types of electrodes and a few examples of the kinds of potentials they record will be given here. *Macroelectrodes*, which are large in size in comparison with a single neuron or receptor cell, record the spatially summed potentials from many cells. In or near the cochlea such electrodes pick up two classes of potentials, one known as the *cochlear microphonic* (*CM*), which represents the spatially summed sensory receptor potentials of the activated hair cells and which also roughly reproduces the stimulus waveform. The other potential, originally designated N_1 or the "first neural" potential, represents the spatially summed, temporally synchronized impulses in auditory nerve fibers in response to a transient signal such as a click (see Chapter 108). In general, recordings with macroelectrodes reveal synchronized neural discharges generally referred to as *evoked potentials*. Evoked potentials at the level of the brainstem are utilized in the clinic to evaluate the function of the middle ear and cochlea in patients. In experimental animals such recordings can be used to "map" auditory areas at many levels of the nervous system and to study the organization of the responsive neural fields, including what is referred to as *tonotopicity*, which is the orderly representation of stimulus frequency along one or more axes within the field.

Detailed studies of *stimulus coding* in the discharges of a single neuron became possible with the introduction of *extracellular microelectrodes*, which have tips small enough to pick up the external field potentials from only one cell body or nerve fiber. The propagated trains of all-or-none discharges in a single auditory nerve fiber in response to varying stimuli reveal how the attributes of a stimulus (e.g., frequency, level, duration) are represented in the activity of the fiber. The *temporal patterns* of discharges in the entire array of fibers activated by a stimulus, together with the *spatial pattern* of cochlear locations (or *places*) from which the activated fibers project, represents all of the information about that stimulus that is available to the central auditory nervous system (see Chapter 109). Similar studies of single neuron discharges at higher levels of the auditory system reveal how more complex stimulus attributes are encoded, such as the binaural interaction observed with differences in intensity and arrival time of sounds at the two ears; such interaction is believed to be related to mechanisms involved in sound localization (see Chapters 110, 111, and 113).

While studies of stimulus coding are typically carried out on single neurons, the investigator generally attempts to record from a large enough sample of cells to estimate the behavior of the whole population of similar neurons; such studies thus represent a kind of *systems analysis* with the objective of learning how the entire system of related neurons behaves. In contrast, studies of individual neurons using microelectrodes small enough to penetrate the cell membrane without causing significant alteration in cellular function yield so-called *intracellular recordings*. Such recordings can reveal many details of cellular function, including *excitatory and inhibitory postsynaptic potentials*, which, together with the intrinsic properties of the cell membrane, determine how the neuron *processes* incoming information before releasing *neurotransmitter* in synaptic transmission to the next cell or cells in the neural circuit. Continuing advances in molecular biology and biophysics are yielding fundamental new insights into membrane physiology through development of the technique of *patch-clamping* in which an electrode is sealed to an area of cell membrane so tiny that movements of ions into and out of the cell in a single channel can be recorded while the membrane potential is held at any desired value. Among many other possibilities, this technique holds promise for increased understanding of the gating mechanisms that control the flow of ions in hair cell mechano-transduction.

In summary, electrophysiological studies constitute one of the most important sources of information on the function of the auditory system, whether that function be the mechanisms by which a single hair cell in the cochlea responds to movement of its stereocilia or the behavior of the vast population of neurons that underlie sound localization.

4.3 Imaging

From the very beginning of research in PA, images of structure have played a vital role, starting with the first sketches of the gross anatomy of the ear, continuing with light microscopy and its depiction of cellular groupings and their interconnections, followed by transmission electron microscopy (TEM) and the ultrastructure it reveals and culminating in scanning electron microscopy (SEM), which yields such striking three-dimensional views of hair cells with their orderly patterns of stereocilia. The principles of systems analysis clearly require identification of the components of the system and their interconnections while cellular and molecular studies of single cells of all varieties are critically dependent upon knowledge of their ultrastructure. It is for these reasons that neurophysiologists and neuroanatomists have long recognized the rewards to be gained by closely correlated studies of structure and function. Today that correlation is often implemented by the injection of dye via an intracellular recording microelectrode into a single neuron whose function has been studied earlier. Subsequent to appropriate histological staining of the tissue, all of the processes of the cell, including cell body, dendritic tree, and axon, will be clearly marked and may be traced for entry into a computer, together with a record of the cell's response properties. When a collection of similar cells has accumulated, it is frequently possible to draw conclusions that relate structure and function. This approach has been applied to several of the auditory nuclei, especially the cochlear nucleus (CN), which is the first way-station along the auditory projection pathway (see Chapters 110 and 111). Several discrete categories of temporal discharge patterns have been identified in CN neurons, and these patterns have been found to correlate well with cell types defined on the basis of morphology.

Another kind of imaging tool that is experiencing steady growth in usage in PA research is nuclear magnetic resonance imaging (NMRI), which produces images of brain sections that show localized areas of change in brain activity in response to sensory stimuli or tasks such as mental arithmetic. Since the technique is noninvasive and easily tolerated by human subjects, NMRI has the potential to provide much more straightforward data on the localization of various classes of brain function than the methods currently in use.

4.4 Additional Methods

It is obvious that the methods outlined above represent but a fraction of the techniques currently in use in auditory neuroscience. One other area of increasing importance is described in Chapter 112 on the biochemistry and pharmacology of the auditory system. Much progress has been made in identifying neurotransmitters that are the chemical *messengers* through which neurons communicate at synapses. Yet another area of great current interest is that of developmental studies of the nervous system, including the question of regeneration of damaged or destroyed sensory receptor cells as well as the role of genetic linkages in normal and abnormal development. Lastly, two methods that increase access to individual cells under direct visualization, improve mechanical stability for recording purposes, and control the biochemical and pharmacological milieu are tissue culture and slice preparations of brain tissue, both of which are enjoying increasing usage.

5 THE "ACTIVE" COCHLEA

The remarkable capabilities of the auditory system continue to pose intriguing challenges to our understanding. For example, at the threshold of hearing, human listeners can detect sounds that cause displacements in the cochlear sensory organ that are subatomic in size, of the same order, or less than the displacements calculated to arise from thermal noise agitation. Full details of the transduction process, in which mechanical displacement of the cilia of receptor cells triggers impulses in afferent nerve fibers, remain an elusive goal, central to the understanding of auditory function. Another capability that has long puzzled theorists is the capacity of listeners to discriminate changes in the frequency of a pure tone of less than 0.5% for which any known design of electrical filter of comparable bandwidth would have a response time-constant many times greater than the time required for the auditory system to discriminate differences in the arrival time of a sound at the two ears which may be as little as 10 μs.

Stimulated by such puzzles and fueled by rapid advances in neurobiology over the past two decades, the pace of research on cochlear function has steadily quickened. In the overall scheme of sensory systems, hearing is considered to be one of the *special senses* or *modalities*, so designated because it involves highly specialized mechano-receptors known as *hair cells*, which are activated by bending of their *hair bundles* or *ciliary tufts*. Hair cells are found in a wide variety of sensory organs ranging from the *lateral line* organ in fish to the *inner ear* of mammals. The latter contains six different sensory organs. Five of the six encode head-tilt and linear and angular acceleration, information essential to the maintenance of balance. The sixth is the so-called *organ of Corti* in the fluid-filled, snail-shaped *cochlea* that contains two kinds of hair cells, the *outer hair cells* (*OHCs*) and the *inner hair cells* (*IHCs*). Omitting many important details, the OHCs receive information from the central nervous system (CNS) via *efferent* nerve fibers while the IHCs are innervated by the vast majority of *afferent* fibers that transmit information about an auditory stimulus to the CNS (see Chapter 107 for details of cochlear anatomy).

Throughout the cochlear spiral, the organ of Corti is supported by the *basilar membrane* and covered, in part, by the overlying *tectorial membrane*. The modern era of understanding of mechanical events in the cochlea originated over half a century ago with the Nobel prize-winning work of Georg von Bekesy[1]; he showed that vibration of the *stapes* in the *oval window* generates sound pressure in the cochlear fluid, causing the basilar membrane to undergo a *traveling wave* of deformation that begins at the *base* and proceeds toward the *apex*. The mechanical tuning of the basilar membrane varies systematically from high frequencies at the base to low frequencies at the apex, primarily due to a decrease in stiffness from base to apex, resulting in a frequency-to-place mapping along the cochlear spiral. The cochlea thus functions as a kind of hydromechanical frequency analyzer in which the mechanical input to the hair cells varies according to their locus or place. Afferent nerve fibers innervating the basal IHCs have tuning curves with high *best* or *characteristic frequencies* (CFs) while those closer to the apex exhibit correspondingly lower CFs.

Until about two decades ago it was almost universally believed that cochlear hair cells function as *passive* transducers that undergo depolarization upon mechanical deformation of their stereocilia. This causes release of excitatory neurotransmitter at synapses with afferent auditory nerve fiber endings, triggering propagated all-or-none impulses in the fibers. In the early 1970s evidence began to accumulate suggesting that the stereocilia of some types of hair cells might exhibit motility so that the cell could function as both a sensor and a motor element. The initial evidence was anatomical, showing that some stereocilia contain actin, a component of muscle fibers associated with their contractile mechanisms. Later, visual observation of isolated hair cells that had been removed from living animals confirmed that the cells did indeed exhibit motile behavior when subjected to electrical stimulation. Continuing investigation and theorizing gave rise to the concept of the "*active*" cochlea[2] with increased sensitivity and greater sharpness of tuning in comparison with the passive models previously accepted (see Chapters 107 and 108). Although differing widely in detail, many hypotheses have been advanced which assume that the OHCs in the active cochlea change their mechanical properties under control of the CNS via the efferent projections, thereby functioning as a feedback system.

Major advances in our understanding of the details of transduction have resulted from remarkably delicate experiments on hair cells in lower forms. Changes in the intracellular potential of hair cells from the bullfrog ampulla have been directly measured as their hair bundles were displaced.[3] Another series of fundamental studies on hair cells in the turtle cochlea has shown that the mechanical frequency tuning exhibited by the cell is determined by resonance of the equivalent electrical circuit of the cell membrane impedance rather than from mechanical resonance of the surrounding structures as is the case in mammalian cochleas.[4]

While hair cell physiology is yielding radically new concepts about function, investigators are developing increasingly sensitive instruments to measure basilar membrane vibration, including refinements in the Mössbauer technique and laser interferometry (see Chapter 108). Whereas Bekesy, working in cadaver cochleas at the very high stimulus levels required by his optical measurement technique, had reported linear basilar membrane vibration and broad tuning, recent studies demonstrated that basilar membrane vibration near the best frequency at the point of observation is *nonlinear* and that mechanical tuning could be as *sharp* as the tuning curves of single auditory nerve fibers, provided the cochlea was in good physiological condition and only moderate stimulus levels were used. As the state of the cochlea deteriorates or when higher stimulus levels are presented, the tuning curves broaden and the system becomes linear, similar to Bekesy's observations. Thus most students of cochlear function now postulate that some kind of active feedback mechanism mediated by the OHCs is superimposed on the passive cochlear behavior described by Bekesy. The result of this active mechanism is an increase in sensitivity of as much as 40 dB, together with substantial sharpening of the tuning curves of auditory nerve fibers.

The contractile behavior of hair cells in active cochleas has also provided one basis for explaining the origin of so-called otoacoustic emissions, which are sounds *generated in the cochlea* and transmitted *backward* via the ossicles and tympanic membrane *to the outer ear canal*, presumably because excessive positive feedback results in oscillation (see Chapter 108).

6 RELATION TO OTHER FIELDS IN ACOUSTICS

If attention shifts from physiological activity to *sensation* and *perception*, a disciplinary boundary is crossed into the realm of psychological acoustics (see Part XII), another subdivision of acoustics closely related to PA. The border between physiological and psychological acoustics has long been fertile ground for interdisciplinary interaction; the findings on one side of the boundary often stimulate and guide research and theoretical interpretation on the other. Although there are important exceptions, up to the present time studies in physiological acoustics have been carried out primarily on subhuman species while psychological experiments have typically utilized human subjects. The ongoing rapid development of noninvasive imaging techniques [such as NMRI and positron emission tomography (PET) scans] holds great promise that, in the future, many experimental observations will be easily performed on humans, thus eliminating the need to assume that the results in an experimental animal apply directly to humans. Physiological acoustics also has ties to other areas of acoustics, including speech processing and perception, musical acoustics, and bioacoustics.

7 STANDARDIZED TERMINOLOGY FOR STIMULUS AND RESPONSE PARAMETERS

Several conventions in terminology have been adopted in both physiological and psychological acoustics in order to minimize ambiguity by distinguishing between stimulus and response properties. *Psychological attributes* of sound include qualities such as *pitch*, *loudness*, and *timbre*, which are used only to describe *sensation* and/or *perception*, that is, "states" or "processes" internal to a listener. *Stimulus attributes* include such physically measurable quantities as *frequency*, *phase*, *sound pressure*, and *spectrum level*. Other stimulus parameters of importance in both physiological and psychological acoustics are specifications concerning the precise manner in which stimuli are applied to the two ears. If a stimulus is presented to only one ear via an earphone or through blockage of the other ear, the stimulus is referred to as *monaural* or *monotic*. *Binaural* means that both ears are stimulated (as distinguished from monaural), but the signals to the two ears may be the same or different. The term *diotic* is used when an identical stimulus is presented to both ears (usually via earphones), while *dichotic* describes the situation in which the stimulus to each ear is different and individually controlled (again via earphones).

Not infrequently, authors in both the physiological and psychoacoustic literature are guilty of imprecise use of the term *intensity* to describe the magnitude of an auditory stimulus. With a few important exceptions (see Chapters 105 and 106), acoustic intensity is not often actually measured because of the complexity of the measurement technique. Instead, most investigators use signal-generating equipment that controls the electrical input to a transducer, the acoustic output of which is calibrated in terms of *sound pressure* (linear scale) or *sound pressure level* (log scale). Although calibrated microphones are now widely used in PA, an occasional publication will be encountered in which stimulus strength is expressed in *decibels of attenuation* below the maximum available output level, which is often not specified. Another way to describe the relative magnitudes of stimuli is to indicate the amount (usually in decibels) by which the level of a specified stimulus exceeds that of a reference stimulus; when the reference stimulus is the threshold of a behavioral response, the magnitude of the specified stimulus is expressed as *dB SL* (*sensation level*).

8 COMMENTARY ON CHAPTERS IN PART XI

1. *Chapter 104: Auditory Function.* Differing from all but one of the other chapters in this part, each of which focuses on a particular aspect of information about the auditory system, Chapter 104 takes a broader view, incorporating findings from many lines of inquiry to yield a coherent picture of the numerous modes in which hearing contributes to daily life. The chapter enters the age-old debate on the relative importance of place and time cues in a variety of discrimination tasks and comes down strongly in favor of a system that utilizes both. The coding of speech sounds is also discussed, including the relative merits of place and time coding in the presence of a masking sound. Although some information about cochlear anatomy and physiology is included, readers lacking a background in the characteristics of the auditory periphery would benefit by first reading Chapters 107–109.

2. *Chapter 105: Acoustical Characteristics of the Outer Ear.* The anatomy of the outer ear is described in detail, together with a rigorous analysis of the acoustical properties resulting from its structure. The focus is primarily on the human ear although information on other species is included. The transformation of sound pressure from the free field in the absence of a listener to the pressure measured at the eardrum with the listener in the field is described and interpreted in terms of acoustical theory; this transform varies with frequency and source direction and thus provides cues that could be of use in sound localization. The overall efficiency of the human outer ear as a collector of sound energy is estimated in terms of the diffuse field absorption cross section for a model ear, the values of which are close to the theoretical limit for frequencies above about 2.5 kHz. The properties of some artificial ears, earphone couplers and ear simulators are also discussed. The chapter should be of interest to readers seeking a general understanding of the role of the external ear in hearing as well as to specialists in need of detailed, quantitative information on this structure.

3. *Chapter 106: Acoustic Properties of the Middle Ear.* The chapter begins with a thorough description of the structures in the mammalian middle ear, followed by an equally thorough discussion of their functions. Starting with the vibratory representation of an auditory stimulus in the motion of the eardrum and the attached malleus, the discussion follows the conduction of stimulus energy to the movement of the stapes footplate in the oval window of the cochlea, which creates sound pressure in the perilymphatic fluid. The transfer function of the middle ear is modeled as an acoustic two-port network involving the acoustic impedance of the eardrum and that of the stapes footplate, both quantities having been measured in humans and in several experimental animals. The discussion considers alterations in middle ear transmission due to ear pathology, to changes in static pressure in the middle ear cavity, and to contractions of the middle ear muscles. Under normal circumstances and in the absence of muscle contractions, the middle ear operates as a linear system up to very high sound pressures. The ratio of sound pressure measured in the cochlear fluid to that measured in the air in the outer ear canal near the eardrum was found to be between 20 and 30 dB for frequencies from 1 to 8 kHz for cat, guinea pig, and chinchilla; these values are consistent with the ratio of the area of the eardrum to that of the stapes footplate in each of these species, showing that most of the pressure gain produced in the middle ear is due to this area ratio.

4. *Chapter 107: Anatomy of the Cochlea and Auditory Nerve.* This chapter presents a detailed description of the morphology of the cochlea and its innervation. Additionally, interwoven into the stream of structural information is an equally detailed description of function at a qualitative level. The result is a concise, coherent blend of cochlear anatomy and physiology that should be of interest to both novices and experts. It is recommended that readers lacking a background in PA read Chapter 107 before consulting other chapters in Part XI.

5. *Chapter 108: Cochlear Mechanics and Biophysics.* This chapter presents a comprehensive review of cochlear function, from the passive, linear mechanics described by Bekesy for cadaver ears and intense stimuli to the current concept of the active cochlea with its physiologically vulnerable feedback system involving outer hair cell motility. The coverage is very broad, including information on hair cell transduction, cochlear potentials, and otoacoustic emissions in addition to basilar membrane vibration. The effects of different kinds of

stimuli are also considered, together with the generation of distortion products. Sufficient information on cochlear anatomy is provided that the chapter can be consulted on a stand-alone basis for readers having a background in this topic.

6. *Chapter 109: Physiology of the Auditory Nerve.* This chapter reviews the characteristics of the trains of all-or-none discharges that may be recorded in single afferent fibers of the auditory nerve and in single efferent nerve fibers in the olivocochlear bundle. The former convey information about an auditory stimulus from the cochlea to the CNS while the latter transmit information from the central nervous system to the outer hair cells in the cochlea. Auditory nerve fibers are characterized by their rates of spontaneous discharges (in the absence of an overt stimulus), their tuning properties, and the extent to which their discharges are time-locked to features in the stimulus waveform. Investigations of tuning have typically involved plots of discharge rate or threshold level vs. frequency for one fiber. This chapter describes an alternative measure in the form of a "population response profile" for which discharge rates for a large number of fibers to the same stimulus are recorded sequentially and then plotted vs. the characteristic or best frequency of the fibers; by means of the known relationship between best frequency and distance along the cochlear partition, the discharge rates of the collection of fibers are plotted vs. the calculated place of innervation for each fiber. Such profiles show the spread of excitation in the cochlea as stimulus intensity is raised; they also reveal differences in the responses of fibers that have different rates of spontaneous discharge. Consideration is also given to the coding of intensity in afferent fibers, together with nonlinear response characteristics and responses to speech sounds. Information is also provided on the response properties of efferent fibers that, unlike afferent fibers, may respond to stimuli applied to both ears; the efferents also exhibit much longer latencies (time from stimulus to response) than afferent fibers. The chapter includes a large amount of information that will probably require a fair amount of background for full appreciation of its content.

7. *Chapter 110: Anatomy of the Central Auditory Nervous System.* This chapter provides a concise but thorough description of all of the well-known landmarks in the central auditory system. However, although the anatomical information is very clearly presented, what makes this chapter especially valuable to hearing scientists who may be only tangentially interested in anatomy, is the continuing insertion of information on relevant auditory function. Such an interleaving of structural and functional concepts facilitates the formation of cross-disciplinary concepts and makes for most interesting reading. The chapter also makes contact with the field of computational neuroscience through the incorporation of ideas such as parallel processing and feedback loops. The chapter provides a good introduction to the anatomy of the central auditory nervous system for novices, but it is also likely to appeal to more senior scholars as a refresher.

8. *Chapter 111: Electrophysiology of the Central Auditory Nervous System.* This chapter reports the results of an extensive survey of the literature on electrophysiological studies at all levels of the central auditory system at which recordings of single units have been made, from the cochlear nucleus in the brainstem to the auditory cortex. Among the attributes of auditory stimuli that have been investigated are frequency, which is represented at all levels of the brain in the form of tonotopic organization along at least one axis in a field of responsive neurons. Responses to interaural level differences (ILDs) and interaural time differences (ITDs) are usually investigated in neurons having inputs from the two ears; such cells often exhibit marked binaural interaction and are thus believed to be involved in sound localization. Information is also presented on the different morphological types of cells that are encountered in the various nuclei. Many cell types classified on the basis of morphology can be identified by their unique temporal discharge patterns. Table 1 of the chapter summarizes this relationship for cells in the CN. Sufficient anatomical information is included that the chapter can probably function on a stand-alone basis for many readers.

9. *Chapter 112: Biochemistry and Pharmacology of the Auditory System.* The volume of information required to present anything approaching an in-depth review of these two fields could easily exceed the entire space allotted to PA in this handbook. Through the use of tables and a concise style of writing, the chapter covers a surprisingly broad range of topics, including the chemical composition of cochlear fluids, cochlear metabolism and drugs that affect it, and neurotransmitters in the cochlea and cochlear nucleus. Throughout the chapter the discussion includes frequent references to the physiological function of the various structures in hearing, both in normal and drug-affected states, thus making the material more interesting and easier to follow for readers with training in hearing science but lacking a background in biochemistry and pharmacology. A comprehensive list of references is included.

10. *Chapter 113: Parallel-Hierarchical Processing of Complex Sounds for Specialized Auditory Function.* As indicated earlier, this chapter differs in style and objective from most of the other chapters in Part XI, which are intended primarily to be sources of information on relatively circumscribed topics in PA. In contrast, this chap-

ter is presented to illustrate the sophistication with which research at the cutting edge of the field is being conceived and conducted. The subdivision of bat auditory cortex into tiny functionally defined areas as described in this chapter is very likely the most detailed description of auditory cortex organization ever observed in any species. The mapping of cortical organization became possible only after lengthy analysis of the emitted cries and reflected echoes in prey-seeking behavior, both in real life and in the laboratory. The success of this investigation clearly demonstrates the advantage of selecting as experimental subjects animals that exhibit highly specialized behavior. It is also interesting to note that the organization of the bat nervous system to carry out the necessary processing is referred to in terms that originated in computer design but have now become part of the vocabulary in computational neuroscience.

REFERENCES

1. G. von Bekesy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
2. P. Dallos, "The Active Cochlea," *J. Neurosci.*, Vol. 12, 1992, pp. 4575–4585.
3. A. J. Hudspeth, "How the Ear's Works Work," *Nature*, Vol. 341, 1989, pp. 397–404.
4. R. Fettiplace, "Electrical Tuning of Hair Cells in the Inner Ear," *Trends Neurosci.*, Vol. 10, 1987, pp. 421–425.

104

AUDITORY FUNCTION

STEVEN GREENBERG

1 INTRODUCTION

Through complex physiological processes, still poorly understood, the auditory system provides an *acoustical image* of the external world. It accomplishes this function by detecting and localizing external sounds, separating sound sources from each other, creating coherent representations of objects and events from the acoustic stream and performing a frequency analysis of incoming sound. Among human listeners the hearing sense assumes even greater significance, as it is the primary medium for social communication and interaction.

Traditionally, the role of the auditory periphery has been viewed primarily as a frequency analyzer, providing a faithful representation of the spectrotemporal properties of the acoustic waveform to central auditory neural populations for sophisticated feature extraction. In recent years this linear, hierarchical perspective has given way to a more complex, decentralized, and distributed paradigm of hearing optimized for processing sound under a wide, often unpredictable range of acoustical conditions that occur in everyday life.

This chapter discusses some of the physiological mechanisms that enable the auditory system to process complex acoustic signals under real-world conditions where reverberation, intense background noise, and competing sound sources commonly occur. Particular emphasis is placed on the means by which the auditory system creates robust representations of communicationally significant information through the use of distributed, multiple representations and coding strategies. Speech will form a primary focus of the discussion because its properties illustrate so well the nature of the nonlinearities and distributed coding inherent in the auditory processing of sound.

2 BACKGROUND

Our ability to interact with the external world depends to a large degree on our acoustical sense. Imagine a world without sound, where objects are seen, touched, and smelled but not heard. Although our reliance on the hearing mechanism goes largely unnoticed in our daily lives, the experience of those who have actually undergone sustained acoustical deprivation indicates that hearing is critical for interacting with the physical environment and maintaining social links to society at large.

The auditory periphery has long been thought to function primarily as a frequency analyzer[1–4] while more sophisticated functions such as complex feature extraction and acoustical recognition have been considered the province of the central pathway.[5]

However, it has become increasingly clear that this traditional view requires serious revision in order to account for the auditory system's capability of encoding meaningful information under a wide range of acoustical conditions. Far more appears to be involved in analyzing sound than merely computing a running spectrum. Under noisy conditions, a faithful representation of the spectrum may even hinder the ability to understand the speech signal as a consequence of acoustic interference, such as background noise or competing speech. It is likely that the auditory system uses special-purpose strategies to focus on those elements of the acoustic signal most likely to contain the meaningful components that bear only a tangential relationship to a Fourier-like analysis.

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

Traditionally, discussions of auditory function have been formulated in terms of *place*, based on spatially discrete activation of neural elements performing a Fourier-like analysis of the signal spectrum (Ohm's "acoustic law"), versus *time*, pertaining to the temporal statistics of neural discharge and their correlation with the input signal. Over the past two decades an increasing appreciation for the complexity and multiplicity of auditory coding has arisen in the scientific community, so that in the 1990s it has become more common to pose the issue in terms of specifying the multiple representations that apply and the conditions under which each may operate or dominate. The traditional perspectives are briefly reviewed, before turning our attention to more contemporary perspectives.

2.1 Place Principle of Frequency Coding

This place principle of frequency coding has, until recently, dominated theoretical discussions of the mechanisms underlying the neural representation of spectral information. However, it is becoming increasingly clear that this framework cannot, by itself, account for all the ways in which auditory information is processed and assembled into complex representations of the external acoustical environment.

2.2 Temporal Coding

As early as 1843, Ohm's acoustic law was called into question by Seebeck's[6] elegant demonstration that the pitch of a complex tone, composed of many harmonics, remains unchanged when the fundamental (lowest) component is absent (canonized as "the case of the missing fundamental"). This demonstration sparked interest in alternatives to Ohm's[1] (and subsequently Helmholtz's[2]) place model. The most popular alternative was a model in which temporal properties of the waveform would be encoded in the discharge of auditory neurons.[7,8] However, such temporal models lay largely out of favor until a decade ago. Several factors have swung the place–time pendulum back toward the temporal domain.

2.3 Limited Dynamic Range of Auditory Neurons

Listeners are able to accurately distinguish and discriminate sounds over a sound pressure level range of 100 dB, yet most auditory-nerve fibers increase their firing rate only over a 20–30-dB range of intensities. This *dynamic range problem* has vexed and intrigued the auditory community for many years[9] but so far has eluded a simple solution. Although there exist powerful mechanisms, such as lateral suppression and inhibition[10] that partially compensate, the fact remains that most neurons in the auditory nerve are discharging at their maximum (saturation) rate in response to a large range of moderate and high-intensity acoustic signals,[11,12] as is discussed further in Chapter 109.

In consequence, intense, broad-bandwidth signals, such as speech, will evoke a pattern of activity across the auditory nerve that exhibits little differentiation in terms of the excitation level of individual neural elements, regardless of the signal's spectral profile. Much of the information in the speech signal is contained within the distribution of energy across the frequency spectrum. For example, the vowel [a] (as in "hot") is spectrally distinguished from [I] (as in "hit") by virtue of its primary locus of energy (formant) at around 1 kHz. In contrast, [I] contains most of its energy around 400 Hz, with a second maximum around 2000 Hz. A strictly place representation of the vocalic spectrum would be expected to contain some form of distinguishing marker in the neural activity pattern in the auditory-nerve fiber array most sensitive to 1 kHz in response to [a] and a corresponding marker around the 400-Hz location for [I]. However, the place-rate representations pertaining to the two vowels are very similar across the auditory nerve, with the exception of a small proportion (about 15%) of fibers with very low background firing rate (see Chapter 109 for further discussion).

Such results have naturally lead to the search for other features in the discharge behavior of auditory neurons that might encode such important features of the spectrum more faithfully. Young and Sachs[13] were among the first to show that perceptually relevant attributes of the vocalic spectrum are encoded in terms of the temporal properties of auditory nerve fibers. Although the manner in which they choose to represent the temporal component of the neural response is controversial, there is little doubt that a large proportion of auditory-nerve fibers exhibit robust phase locking to the perceptually significant lower formants of vocalic signals.

2.4 Auditory Scene Analysis

Another important factor motivating a reevaluation of the relative contributions of place and temporal mechanisms is our ability to distinguish and segregate sounds emanating from a variety of sources, such as different musical instruments or speakers. It is difficult to conceive how the ensemble of frequencies associated with a complex acoustic event, such as a string quartet, could be encoded entirely on the basis of place cues—there are just too many frequency components to track through time. In a manner yet poorly understood, the auditory system utilizes efficient parsing strategies to encode not only information pertaining to a sound's spectrum but also tracks that signal's acoustic trajectory through time and space, grouping neural activity into singular acoustic events attached to specific sound sources.[14,15]

Although the neural underpinnings of auditory scene analysis have been little studied, there is increasing evidence that temporal mechanisms play an important role. Neural discharge synchronized to specific properties of the acoustic signal, such as the waveform modulation frequency (often, but not always, correlated with the signal's pitch) and signal onsets, can function to mark activity as coming from the same source. The operational assumption is that the auditory system, like other sensory systems, has evolved to focus on acoustic events rather than merely performing a frequency analysis of the incoming sound stream. Such relevant signatures of biologically relevant events include common onsets and offsets, coherent modulation, and spectral trajectories.[14,16] In other words, the auditory system performs intelligent processing on the oncoming sound stream to recreate as best it can the physical scenario from which the sound emanates. In the context of a musical performance such scene analysis might involve reconstituting the individual instruments of a string quartet or focusing on the piano solo in a concerto. An auditory scene analysis of Yosemite Valley might entail filtering out the background roar of the waterfalls in order to hear the cry of a red-tailed hawk flying high above or a companion's verbal comment. This ecological acoustical approach to auditory function stems from the pioneering work of Gibson,[17] who considered the senses as intelligent computational resources designed to recreate the external physical world.

Within the traditional framework of hearing theory incoming sound is passively analyzed according to principles established by the mechanical properties of the middle and inner ears, coupled to the physiological properties of the auditory periphery and more central regions of the auditory pathway (see Chapters 108 and 109). The representation (and interpretation) of any sound is therefore largely determined by inherent properties of the sensory system. What we hear is largely determined by the constraints imposed by the anatomy and physiology of the auditory pathway. Emphasis is placed on limits imposed by the system for providing a faithful rendition of the external world. Difference limens for frequency, intensity, localization, and so forth are interpreted as the result of limitations in the construction and operation of the auditory pathway.

The Gibsonian perspective, on the other hand, emphasizes the deductive capabilities of the senses to infer the conditions behind the sound, utilizing whatever cues are at hand. The limits of hearing capability are ascribed to functional properties of interacting with the environment. Sensory systems need not be any more sensitive or discriminating than required to meet the demands of the external world. Evolutionary processes have assured that the auditory system works sufficiently well under most conditions.

Although there are merits to both approaches, neither really addresses the important evolutionary and developmental issues, particularly as to how the auditory system interacts with other sensory systems and with the cortical regions of the brain. Nor do they provide for a dynamic perspective in which the auditory system is able to modify its representation and interpretation of incoming sounds through learning or efferent control.

3 INFORMATION CODING

Through the auditory nerve flows all of the neural information available to higher auditory centers pertaining to the acoustical environment. In the inner ear of the human are approximately 30,000 auditory-nerve fibers and 3,000 inner hair cells (each of which is innervated by 8–11 fibers).[18,19] The richness of auditory experience suggests the occurrence of considerable postprocessing beyond this caudal station of the auditory pathway.

A key element of acoustical processing is its real-time character. In vision it is possible to apportion one's attention and focus selectivity, scanning the visual field. In contrast, the ears are always "open," even during sleep, processing the incoming sound stream on the fly, typically without benefit of an additional "look," thus requiring other strategies for selective and focused attention.

One such strategy is to focus the system's processing capabilities on those components of the acoustic signal most likely to contain significant information-bearing elements at a relatively peripheral stage, such as the cochlea or cochlear nuclei.[20] Certain components of the acoustic signal, such as waveform onsets, amplitude modulation, and spectral peaks are most likely to encode biologically relevant information.

3.1 Importance of Onsets

Onsets often denote the occurrence of a motion-related event such as the beginning of vocally initiated speech events, the slam of a door, the backfire of an automobile, and so forth. Such transients rarely last more than a few milliseconds, possess a broad spectrum and hence are highly effective as alerting cues. Their neural correlate is the burst of discharge associated with the beginning of a stimulus of sufficient amplitude. In the auditory nerve the probability of firing is usually far higher during the initial 5–10 ms following signal onset than afterward. Already at the level of the cochlear nucleus are cells whose responses are heavily weighted toward stimulus onset, to the virtual exclusion of the steady-state response. As one ascends the auditory pathway, emphasis on the initial component of the signal becomes even more pronounced, so that at the level of the auditory cortex it is unusual for a cell to respond beyond the initial few milliseconds of the sound.[21] The signal representation has been transformed from a fine-grained analysis at the periphery to one favoring event detection and discrimination at the cortical level.

This physiological bias toward onsets is commensurate with the observation that the initial few tens of milliseconds of a sound are perceptually more important than what follows, even for relatively long-duration signals. This has been demonstrated by showing that the timbre of musical instruments is heavily influenced by a signal's initial 50 ms.[22] A violin tone can be digitally transformed into a trumpet (and vice versa) by manipulating the rise-time properties of this initial interval, while keeping the remainder of the signal constant. Similar effects have been demonstrated for speech signals.[23]

It would thus appear that the auditory system focuses on the initial portion of a signal, performing a fine-grained analysis for the first 50 ms or so, and then "relaxes," on the assumption that what follows is part of the same "event." On this view the perceptually relevant portion of the signal is correlated in time with the distribution of neural activity. The higher cortical centers appear to be primed to "look" where the action (potentials) is and to ignore the remainder. In this sense, neural adaptation may be regarded as a "novelty" detector, acting as a filter for changes in sensory input.

3.2 Amplitude Modulation

The vibration of the vocal folds during phonation, the rumble of a train's engine, a moth's wing beatings all produce cyclic variations in the acoustic waveform that leave a distinctive neural imprint. Such signals can be thought of as containing a succession of onsets, each modulation cycle being treated as a separate event trigger. The vibration of the laryngeal vocal folds of the human voice is on the time scale of milliseconds (3–12 ms). Each glottal pulse produces in the auditory periphery and cochlear nucleus a burst of discharge activity correlated with the amplitude variation of the waveform, indicating the time location of perceptually important information.[24,25] The amplitude modulation associated with speech syllables and segments is on a longer time scale (100–400 ms).

Synchronization to the amplitude modulation characteristics of the waveform is pervasive in the auditory nerve[26,27] and is also common in the cochlear nuclei.[28,29] It is not unusual for cells at this level to synchronize their discharge to frequencies between 100 and 1000 Hz, a range matching that of *low pitch* and the *residue*.[31] At more central stations of the auditory pathway the upper limit of phase locking to AM signals declines, so that at the level of the auditory cortex it is atypical to observe synchrony to modulation frequencies greater than 50 Hz.[32,33] One might consider the auditory periphery performing a fine-grained analysis on a pitch cycle-by-cycle basis while the higher auditory stations are more concerned with encoding longer time span events such as speech syllables and other acoustical objects.

3.3 Noise Reduction

Neural synchrony in the auditory periphery provides an important means of improving the signal-to-noise ratio of the internal representation of signals.[20,34] Consider the situation where speech occurs in background noise. Normal-hearing listeners typically can understand speech even under extremely low (approximately −5 dB) signal-to-noise ratios.[35] This capability is probably a consequence of auditory-nerve fibers and second-order neurons phase locking to the spectral peaks (formants), effectively swamping out the effects of the background noise.

In the cochlear nucleus and higher auditory nuclei, noise filtering is even more effective as a consequence of coincidence detection and neuronal correlational mechanisms. Coincidence detector cells such as the onset-chopper units of the posteroventral cochlear nucleus are more likely to discharge when their auditory-nerve fiber inputs fire within a small (≲100 μs) time window.[30] Such coincident firing patterns are typically produced only for pulselike transients (such as the release associated with stop constants such as [p, t, k]) and for phase-locked responses to amplitude modulated and formant-like signals.[36] Such cells are considerably less responsive to noiselike signals. As such, they act as noise rejection filters effectively cleaning up the auditory representation of the primary signal.

3.4 Reliability of Coding

All sensory systems face coding reliability problems. Because organisms interact with a complex, multifaceted environment, it is difficult to anticipate all of the environmental contexts in which informationally relevant signals will occur. In consequence, coding strategies optimized for automatic suppression of extraneous or interfering signals provide an advantage over more linear forms of processing. It is important to realize that because of this concern for reliability, far more is involved in decoding the speech signal than merely computing a running spectrum. In such noisy conditions, a faithful representation of the spectrum may even hinder the ability to decode or understand as a consequence of acoustic interference. Computing a running spectrum of the speech signal is a singularly inefficient means to accomplish this end, since much of the acoustics is extraneous to the message. The auditory system has developed several strategies for enhancing the reliability of decoding the incoming signal, among the most important of which are lateral suppression and inhibition, neural phase locking and distributed representations of the spectrally most intense regions of the signal.

4 FUNCTIONALLY SIGNIFICANT PROPERTIES OF THE AUDITORY PERIPHERY

Until recently it was generally thought that auditory-nerve fibers were relatively uniform in their response properties.[4] Over the past 15 years it has become clear that there is a wide variation in response properties of neural elements, even at the most peripheral locus of the auditory pathway,[37,38] and it is likely that there are significant functional consequences of this diversity.

4.1 Phase Locking

The modulation of neurotransmitter release by the inner hair cell receptor potential has an important consequence for the coding of complex sounds in the auditory periphery. At low frequencies, below 4–5 kHz, the receptor potential oscillation is sufficiently large to modulate transmitter release in a manner that affects the temporal pattern of auditory-nerve fiber discharge.[39] In response to sinusoidal stimulation, transmitter release occurs only during a restricted portion (the rarefaction phase) of the stimulus cycle. Because the probability of a fiber's discharge is highly correlated with transmitter release, the unit's firing pattern is itself modulated in a fashion analogous to (but differing in important ways from) the inner hair cell receptor potential). The receptor potential modulation is of sufficient magnitude to produce a cadence of discharge activity in auditory-nerve fibers temporally synchronized to the driving waveform (see Chapter 108 for further discussion).

This phase-locked auditory-nerve fiber response provides one means by which frequency is encoded in the auditory periphery. In the absence of phase locking, the probability of firing, relative to the cochlear-filtered waveform, is approximately uniform throughout. Under this condition the firing probability is uncorrelated with the fine temporal structure of the driving signal, thus providing no temporal information with which to infer the spectrotemporal characteristics of the stimulating waveform. This uniform firing distribution is characteristic of the auditory-nerve fiber response to sinusoidal stimuli above 4–5 kHz. Contrast this pattern with that pertaining to lower frequency stimuli. For such signals, the probability of discharge is relatively high during a restricted time interval (or phase) of the stimulus cycle, and relatively low otherwise. Under such conditions potentially important information concerning the spectrotemporal characteristics of the driving waveform are contained in the cadence of neural discharge, as shown in Fig. 1.

4.2 Lateral Suppression

Under certain stimulus conditions, the discharge activity of single fibers actually diminishes in the presence of intense signals. In response to wideband noise, the firing rate of certain fibers, usually of low spontaneous rate, will first grow with increasing stimulus level, and then decrease with further increments of the noise level.[42] A second instance of suppression is in response to the concurrent presentation of two sinusoidal signals. The response (measured in terms of average firing rate) to a sinusoid presented at the fiber's most sensitive (*characteristic*) frequency (*CF tone*) will diminish upon presentation of a second tone (the *suppressor*).[43] The effective frequency and intensity range of the suppressor tone is the suppression region. This area generally lies outside the range of frequencies and intensities effective in driving the fiber to single-tone stimuli (the *excitation* region), as shown in Fig. 2. Under most conditions, the suppressor needs to be at least 20 dB more intense than the characteristic frequency signal to effectively drive down the fiber's discharge rate. It is potentially significant that the spectral magnitude disparities required for suppression are commonplace in many communication signals, including speech. Such suppression also occurs in the auditory nerve in response to vowel sounds.[44]

The origins and functional consequences of this suppression remain obscure and controversial. Much of the suppression appears to be mechanical in origin, reflecting the nonlinear motion of the basilar membrane to complex signals.[10] The functional consequence of this suppression is a sharpening of the tonotopically organized profile of neural excitation.

4.3 Spontaneous Activity

Most auditory-nerve fibers discharge in the absence of external acoustic stimulation. This spontaneous activity is thought to reflect the random release of transmitter from the inner hair cell and its uptake by associated auditory-nerve fibers. Sixty percent of the fibers have spontaneous rates (SR) between 18 and 120 spikes/s (*high-SR* group). A smaller proportion (25%) have SRs between 0.5 and 18 spikes/s (*medium SR*). The remain-

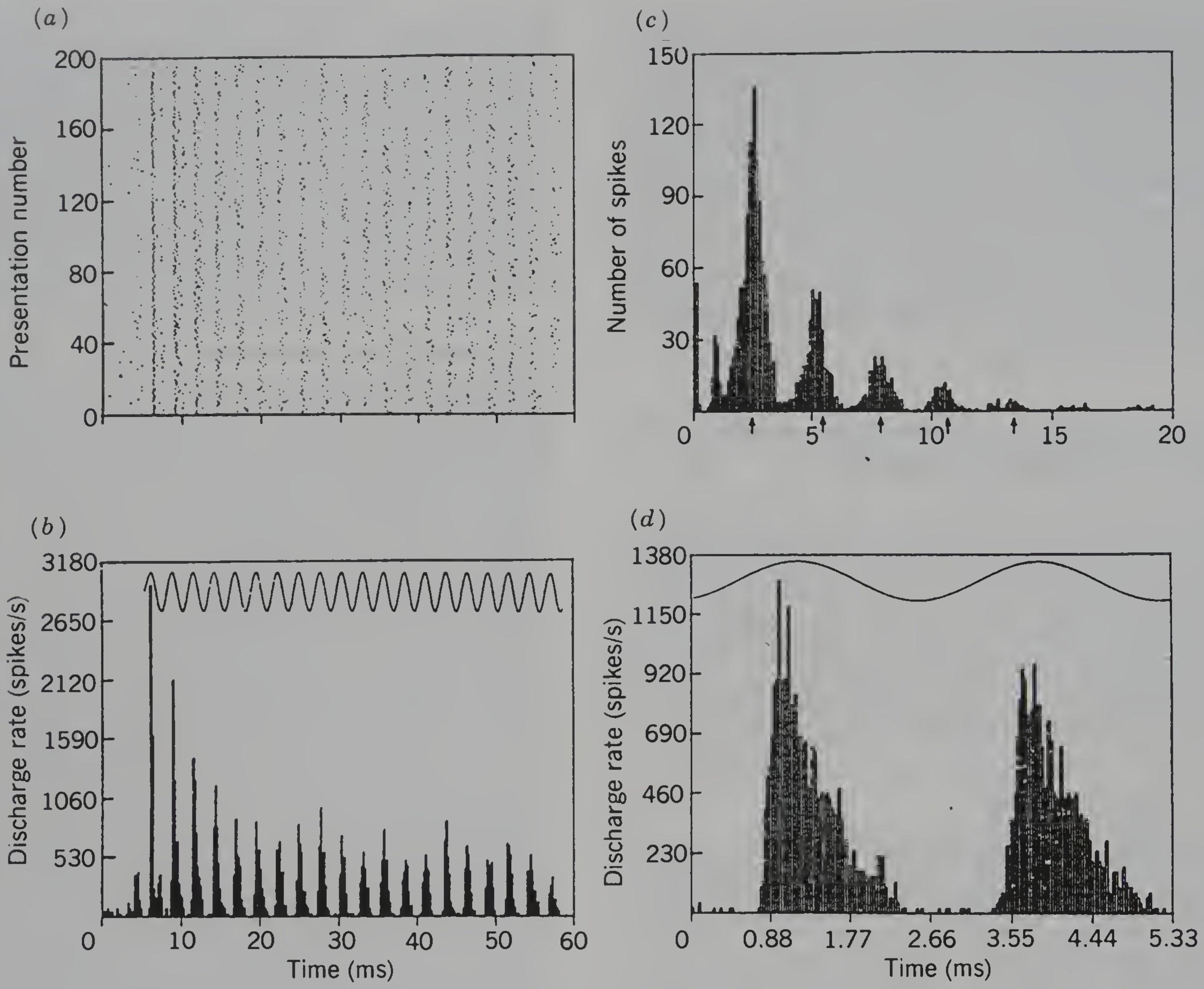


Fig. 1 Phase locking in the auditory periphery. Phase locking of a primarylike unit recorded from the anteroventral cochlear nucleus of the cat, in response to a 375-Hz sinusoid presented at 70 dB SPL. Stimulus duration was 50 ms. The unit's temporal response pattern is similar to that of an auditory nerve fiber. (*a*) Dot raster display (in which each dot represents the occurrence of a discharge) shows the stochastic nature of the phase-locked response. Bin width 250 μs. (*b*) Post-stimulus-time (PST) histogram shows the probability of response as a function of time from stimulus onset. In this instance the probability is indicated in units of instantaneous discharge rate. The average rate was approximately 160 spikes/s. A sinusoidal function, whose period is equivalent to that of the input signal, is shown for reference. Note the rapid adaptation of the response over the initial 5–10 ms. Bin width 250 μs. (*c*) Interval histogram spanning a time window of 20 ms. Arrows mark the stimulus period and integral multiples thereof. Bin width 100 μs. (*d*) Period histogram binned over a time window of 5.33 ms (equivalent to two periods of the input signal). Bin width 26.7 μs. The unit's discharge rate threshold was 25 dB SPL. Its spontaneous rate was 36 spikes/s. Histograms are based on 200 stimulus repetitions. (*e*) The temporal distribution of neural discharge is highly synchronized to the stimulus waveform when the signal frequency is below 2 kHz. At higher frequencies, the correlation between waveform periodicity and neural discharge diminishes appreciably, until it disappears around 5 kHz. Shown are period histograms for a single auditory-nerve fiber in the squirrel monkey in response to sinusoidal signals ranging in frequency from 1 to 5 kHz. [(*a–d*) Reprinted with Ref. 40, p. 8, with permission of Academic Press. (*e*) Reprinted from Ref. 41, with permission of the American Physiological Society.]

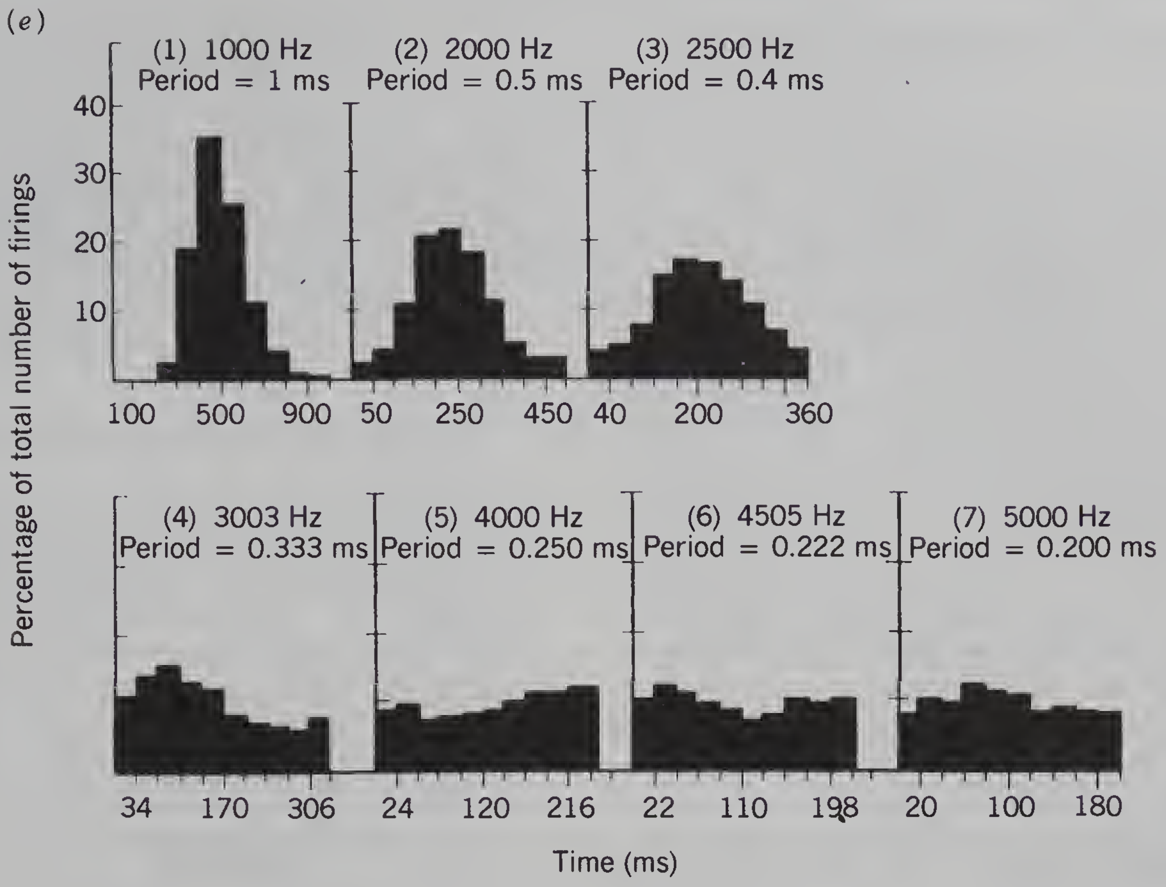


Fig. 1 (*Continued*)

der exhibit little or no spontaneous activity (*low SR*).[37] The underlying basis of these different SR classes remains obscure. However, there are a number of morphological and physiological differences, segregating the three classes, of potential significance for neural representation of complex sounds (see Chapter 109 for further discussion).

Several functional properties distinguish the low- and high-SR groups. These pertain to response threshold, rapid adaptation, phase locking, and suppression.

Response Threshold High-SR fibers are typically 5–10 dB more sensitive than their low-SR counterparts[37,38] when threshold is measured in terms of average discharge rate, with some low-SR fibers as much as 20–30 dB less sensitive. The potential magnitude of threshold disparity is even greater (by 10–15 dB) when the temporal[41] and onset[45] components of the response are taken into account.

Dynamic Range The dynamic range of response (the intensity range over which a fiber increases its firing rate) for low-SR fibers is often two to three times greater than for high-SR units (60 dB vs. 20–30 dB).[46] Much of this extended range occurs at levels above 50 dB sound pressure level (SPL), where the slope of the input–output (I/O) function is relatively low (see Chapter 108 for further discussion). A small subpopulation of fibers in the guinea pig auditory nerve exhibit essentially linear I/O curves over an even larger dynamic range.[47] However, these results have not, as yet, been reported for other mammalian species.

Rapid and Longer Term Adaptation Among high-SR fibers the instantaneous discharge rate is greatest at stimulus onset and declines very quickly over the following 10–15 ms.[45,48] The initial decline is termed *rapid* adaptation and is most pronounced in response to high-amplitude signals. The very lowest SR fibers do not exhibit rapid adaptation, even at high signal intensities.[45] Rapid adaptation for medium-SR units occurs at an intermediate level.[45] Chapter 109 discusses adaptation in further detail.

It has recently been shown that the long-term adaptation time constant of low-SR auditory nerve fibers is substantially longer (several seconds) than that of their high-SR counterparts.[49]

Temporal Coding Another distinction between low- and high-SR fibers concerns their ability to temporally encode low-frequency modulations. The low-SR group is

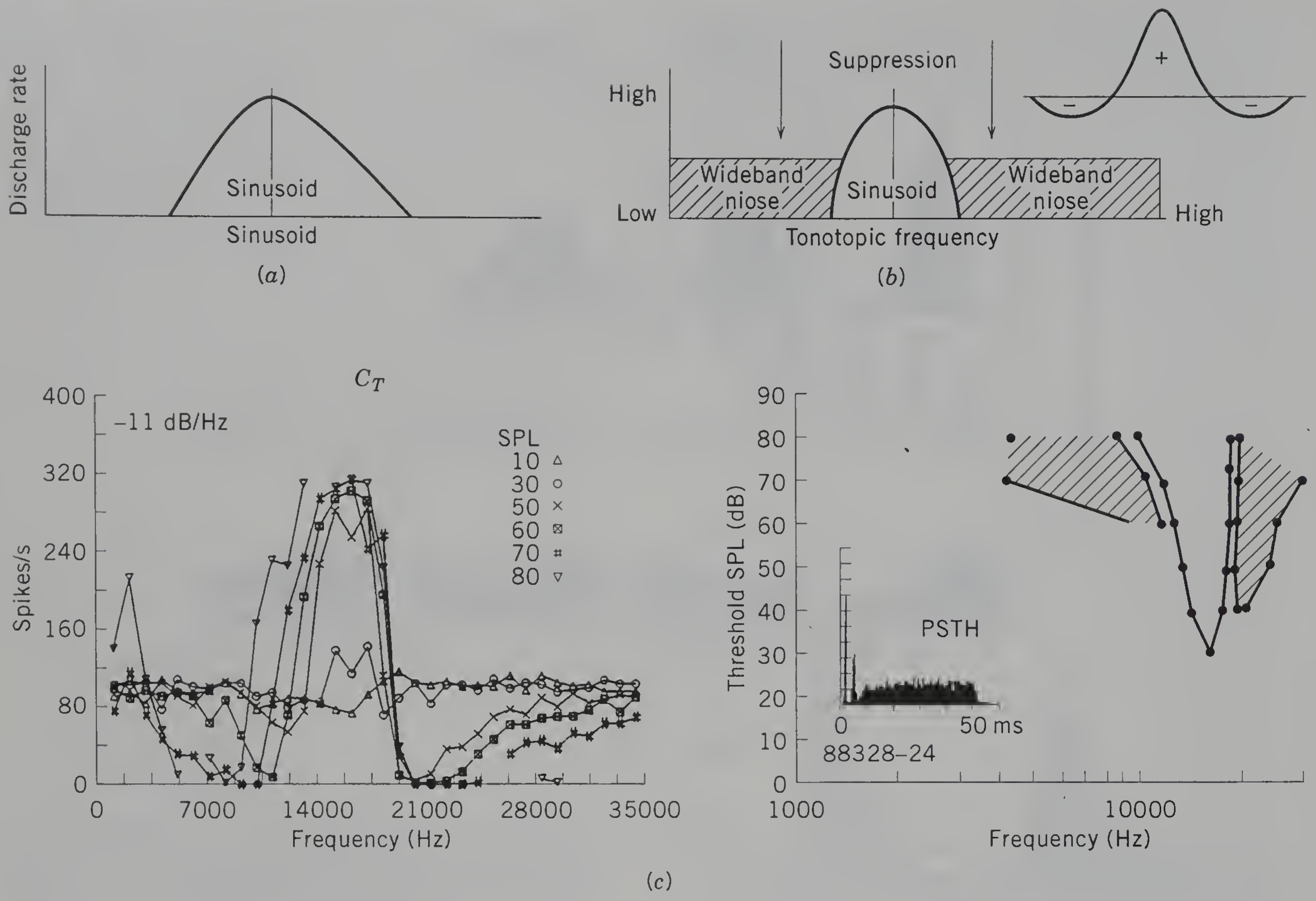


Fig. 2 Lateral suppression and inhibition. Lateral suppression refers to the reduction in discharge rate evoked by a stimulus in the presence of an additional signal, such as a second tone or broadband noise. (*a*) A schematic illustration of lateral suppression and inhibition showing the excitation pattern (stippled area) evoked by a sinusoidal signal (black bar). (*b*) The excitation to the same sinusoidal signal in the presence of broadband noise. The noise effectively sharpens the rate-place representation of the sinusoidal component by suppressing the discharge on the skirts of the excitation pattern. (*c*) Such discharge suppression is shown for a transient chopper unit from the posteroventral cochlear nucleus of the cat. Rather than record from many different cells spanning a broad tonotopic range, the discharge of a single cell is recorded in response to sinusoidal signals of variable frequency and sound pressure level. The degree of suppression is considerable, particularly on the high-frequency side of the excitatory component of the response area. The suppression contour is shown (stippled area) flanking the excitatory region, along with the poststimulus time histogram (PSTH) of the chopper unit. In the present example, the response area delineated by sinusoidal signals is altered in the presence of broadband noise. The changes in the filtering principally affect the edges of the excitatory region, sharpening the contrast between the foreground excitatory signal and the background. This contrast enhancement is clearly shown in the response areas of neurons obtained in the presence of broadband noise. The suppression results in clear delineation of the rate-place excitation pattern evoked by sinusoidal signals in the presence of the noise background. This contrast enhancement is augmented in the cochlear nucleus through inhibitory neural mechanisms. (Reprinted from Ref. 10, p. 500, with permission of the American Physiological Society.)

generally superior in synchronizing its discharge activity to the modulating waveform.[50,51]

Lateral Suppression The low-SR fibers exhibit more rate suppression than high-SR units, suggesting that the low-SR fibers more closely reflect the mechanical motion of the basilar membrane at high sound pressure levels (where the suppression would be most apparent) than their more sensitive high-SR counterparts.[46]

4.4 Significance of Cochlear Nonlinearities

At sound pressure levels characteristic of speech, the response of the cochlear partition is highly nonlinear. The reasons for the nonlinear response are complex and not fully understood. An important source of this nonlinear response is the displacement behavior of the basilar membrane (BM). Above approximately 50 dB SPL, the maximum amplitude of the BM's displacement is compressive, being less than proportional to input energy.[52] In this range a doubling of sound pressure level (i.e., 6 dB) will result in perhaps only a 40% (3-dB) gain in BM displacement magnitude. This nonlinearity is frequency dependent. Thus, the amount of compression is greatest at that BM location being driven the hardest, and is least in regions with relatively low amounts of energy, distant from the compressive segments (see Chapter 108 for further discussion). This frequency-dependent compressive behavior has significant implications for signal processing at moderate and high sound pressure levels because it appreciably broadens the filtering performed by the cochlea.[20,53]

A second source of nonlinearity is in the inner hair cell transduction. Above 50 dB SPL, the behavior of the receptor potential also becomes compressive. Although some of this nonlinearity is a consequence of BM compression, there appears to be an additional mechanism operating to limit the magnitude of the receptor potential.

A third form of nonlinearity limits the rate at which auditory nerve fibers can discharge. Over a 20 to 30-dB range the discharge rate of most fibers can increase from its background (spontaneous) level up to 200–250 spikes/s. With further increases in sound pressure level the fiber's discharge rate fails to increase, having reached its saturation point. The mechanisms underlying rate saturation are poorly understood and highly controversial. Some models[54,55] favor a limitation on the inner hair cell's ability to produce and/or release neurotransmitter as the primary cause (*presynaptic* mechanisms), while others are inclined to weight *postsynaptic* mechanisms, such as neurotransmitter uptake, more heavily. It is also possible that such rate saturation stems from the operation of both mechanisms acting in concert.

5 AUDITORY REPRESENTATIONS OF THE ACOUSTIC SPECTRUM

It is possible to infer the internal auditory representation of spectrally simple and complex signals by recording the response of hundreds of single fibers to the same stimulus. In the typical "population" study the characteristic frequency and spontaneous activity of the fibers recorded are distributed in a fashion thought to accurately reflect the underlying statistics of the relevant auditory structure, most typically the auditory nerve. In this manner it is possible to infer how much information is contained in the distribution of activity across the tonotopic axis to the stimulus spectrum (see Chapter 109 for further discussion).

It is also possible to infer the integrative activity of auditory neurons through computational approaches, by modeling the physical, physiological, and biochemical properties of the relevant auditory structures and constructing a tonotopically organized representation of neural frequency channels. Such models have been developed primarily for the auditory-nerve response to speech sounds[56–59] and complex signals possessing a salient low pitch.[60–62] Computational models have also been applied to the cochlear nuclei[63] and inferior colliculus[64] in response to complex stimuli.

The present discussion focuses on the representation of speech signals in the auditory periphery, where most of the physiological and computational studies have been performed.

5.1 Rate-Place Representation

The population response technique is presumed to provide an indirect estimate of the magnitude and extent of cochlear excitation. Comparison of the population profiles for low- (1-kHz) and high- (8-kHz) frequency signals across a wide range of sound pressure levels indicate a fundamental difference in the spatial distribution of activity. For high-frequency signals the spatial extent of excitation is relatively constant as a function of SPL, indicating that the filtering of that portion of the cochlea remains relatively sharp from low to high intensities. In contrast, the spatial profile associated with the low-frequency signal expands considerably with increasing intensity, so much so that at high SPLs typical of speech (60–80 dB) much of the auditory nerve is responding to this portion of the spectrum (see Chapter 109 for further details).

On the basis of such population studies it is difficult to envision how the spectral profile of vowels and other speech sounds could be accurately and reliably encoded on the basis of rate-place information at all but the lowest stimulus intensities. However, a small proportion of auditory-nerve fibers (15%), with spontaneous rates less than 0.5 spikes/s, may be capable of encoding the spectral envelope on the basis of rate-place information, even at the highest stimulus levels.[65,66] These fibers exhibit extended dynamic ranges and are most sensitive to the mechanical suppression behavior of the basilar membrane.[41,46] Thus, the discharge rate of low-SR fibers, with characteristic frequencies close to the formant peaks, will continue to grow at high sound pressure levels, and the activity of low-SR fibers responsive to the spectral troughs should, in principle, be suppressed by energy associated with the formants. However, such rate suppression also reduces the response to the second and third formants,[43] thereby decreasing the resolution of the spectral peaks in the rate-place profile at higher sound pressure levels. For this reason it is not clear that lateral suppression, by itself, actually functions to provide an adequate rate-place representation of speech and other spectrally complex signals in the auditory nerve. Thus, the case for a rate-place code for vocalic stimuli is equivocal at the level of the auditory nerve. The discharge activity of a large majority of fibers is saturated at these levels in response to vocalic stimuli. Only a small proportion of auditory-nerve fibers resolve the spectral peaks across the entire dynamic range of speech. And the representation provided by these low-SR units is less than ideal, particularly at conversational intensity levels.

The rate-place representation of the spectrum may be enhanced in the cochlear nucleus and higher auditory stations relative to that observed in the auditory nerve. Such enhancement could be a consequence of preferential projection or through the operation of lateral inhibitory networks that sharpen still further the contrast between excitatory and background neural activity.[10]

Many chopper units in the anteroventral cochlear nucleus respond to steady-state vocalic stimuli in a manner similar to that of low-SR auditory-nerve fibers.[66] The rate-place profile of these choppers exhibit clearly delineated peaks at characteristic frequencies corresponding to the lower formant frequencies, even at 75 dB SPL.[66] In principal, a spectral peak would act to suppress the activity of choppers with characteristic frequencies corresponding to less intense energy, thereby enhancing the neural contrast between spectral maxima and minima. It has been suggested that such lateral inhibitory mechanisms may underlie the ability of anteroventral cochlear nucleus choppers to encode the spectral envelope of vocalic stimuli at sound pressure levels well above those at which the average rate of the majority of auditory-nerve fibers saturate.[66]

The evidence is stronger for a rate-place representation of certain consonantal segments. The amplitude of most voiceless consonants is sufficiently low (< 50 dB SPL), as to evade the rate saturation attendant in the coding of vocalic signals. The spectra of plosive bursts, for example, is generally broadband, with several local maxima. Such spectral information is not likely to be temporally encoded due to its brief duration and the lack of sharply defined peaks. Physiological studies have shown that such segments are adequately represented in the rate-place profile of all spontaneous rate groups across the tonotopic axis.[67]

5.2 Phase-Place and Latency-Place Information

In a linear system the phase characteristics of a filter are highly correlated with its amplitude response. On the skirts of the filter, where the amplitude response diminishes quickly, the phase of the output signal also changes rapidly. The phase response, by itself, can thus be used in such a system to infer the properties of the filter.[68] For a nonlinear system, such as the cochlea, phase and latency (group delay) information may provide a more accurate estimate of the underlying filter characteristics than average discharge rate because latency and phase are not necessarily as sensitive to compressive and saturating nonlinearities. Although the average rate of a fiber may not continue to gain with increasing intensity, the unit's phase response may continue to do so, thus providing information concerning the input spectrum beyond what is available in the rate-place population profile at high sound pressure levels.

Several studies suggest that such phase and latency cues occur in the auditory nerve across a very broad range of intensities. A large phase transition is observed in the neural response distributed across auditory-nerve fibers whose characteristic frequencies span the lower tonotopic boundary of a dominant frequency component,[69] indicating that the high-frequency skirt of the cochlear filters is sharply tuned across intensity. A latency shift of the neural response is observed over a small range of fiber characteristic frequencies. The magnitude of the shift can be appreciable, as much as half a cycle of the driving frequency.[69,70] For a 500-Hz signal this latency change would be on the order of 1 ms. Because this phase transition may not be subject to the same nonlinearities that result in discharge-rate saturation, fibers with characteristic frequencies just apical to the place of maximal response can potentially encode a spectral peak in terms of the onset phase across a wide range of intensities.

Interesting variants of this response-latency model have been proposed by Shamma,[58] Deng et al.,[59] and Holton.[25] The phase transition for low-frequency signals should, in principle, occur throughout the entire response, not just at the beginning, as a result of auditory-nerve fibers' phase-locking properties. They propose that such ongoing phase disparities are registered by some form of neural circuitry presumably located in the cochlear nucleus. The output of such networks would magnify activity in those tonotopic regions over which the phase and/or latency changes rapidly through some form of cross-frequency-channel correlation. In Shamma's model, the correlation is performed through the operation of a lateral inhibitory network, which subtracts the auditory-nerve output of adjacent channels. The effect of this cross-channel subtraction is to null out activity for channels with similar phase and latency characteristics, leaving only that portion of the activity pattern where rapid phase transitions occur. The Deng model uses cross-channel correlation (i.e., multiplication) instead of subtraction to locate the response boundaries. Correlation magnifies the activity of channels with similar response patterns and reduces the output of dissimilar adjacent channels. Holton's model looks for large phase derivatives across channels as well as for correlated synchrony across channels. Whether the cross-channel comparison is performed through subtraction, multiplication, or some other operation, the consequence of such neural computation is to provide "pointers" to those tonotopic regions where a boundary occurs that might otherwise be hidden if analyzed solely on the basis of average rate. These pointers could, in principle, act in a manner analogous to "peaks" in the excitation pattern.

5.3 Synchrony-Place Information

Place and temporal models of frequency coding are generally discussed as if they are diametrically opposed perspectives. Traditionally, temporal models have deemphasized tonotopic organization in favor of the fine-temporal structure of the neural response. However, place and temporal coding need not mutually exclude the other. The concept of the "central" spectrum[71,72] attempts to reconcile the two approaches in a unified model of frequency coding. In this model, both place and temporal information are used to construct the peripheral representation of the spectrum. Timing information, as reflected in the interval histogram of auditory nerve fibers, is used to estimate the driving frequency. The model assumes that temporal activity is keyed to the tonotopic frequency representation. In some unspecified way, the system "knows" what sort of temporal activity corresponds to each tonotopic location, analogous to "matched" filters in linear systems theory.

The central spectrum model is the intellectual antecedent of the peripheral representational model of speech proposed by Young and Sachs.[13] As with place schemes in general, spectral frequency is mapped onto tonotopic place (i.e., fiber characteristic frequency), while the amplitude of each frequency is given by the magnitude of the neural response synchronized to that component by nerve fibers whose characteristic frequencies lay within close proximity (one-quarter octave). The resulting *average localized synchronized rate* (ALSR) representation of the stimulus spectrum is illustrated in Fig. 3. The spectral peaks associated with the three lower formants (F_1, F_2, F_3) are clearly delineated in the ALSR representation, in marked contrast to the rate-place representation. This ALSR representation remains relatively stable over large changes in sound pressure level, commensurate with the perceptual constancy of speech information over comparable dynamic ranges.

5.4 Liabilities of Place Models

A principal drawback of place models is their vulnerability to intense background noise, a common occurrence in the real world. Because the frequency parameter is coded through the spatial position of active neural elements, the representation of complex spectra is particularly vulnerable to extraneous interference.[20] Intense noise or competing sounds with significant energy overlapping frequency regions containing primary information have the capability of damaging the auditory representation of the spectrum. This vulnerability of place representations is particularly acute when the neural information is in the form of average rate, as there is no neural marker other than tonotopic affiliation, which carries information about the frequency of the driving signal. Moreover, the perceptual implications of a strictly rate-place model are counterintuitive, for it is implied that the intelligibility of speech should decline with increasing sound pressure level above 40 dB. Above this level the rate-place representation of the vocalic spectrum for most auditory nerve fibers becomes much less well defined, and only the low-SR fiber population continues to encode the spectral envelope with any degree of precision. In actuality, speech intelligibility is somewhat better above 60 dB,[74] where the rate-place representation is not nearly so well delineated.

The ALSR model offers, at first glance, an appealing alternative to rate-based models, in that temporal information is more robust in the presence of noise. The ALSR model assumes the existence of a central mechanism that "knows" the synchrony pattern appropriate to the characteristic frequencies of the projecting fibers, and is capable of filtering out all other temporal activity. However, the ALSR approach requires that at some level of the auditory pathway a correlation be performed

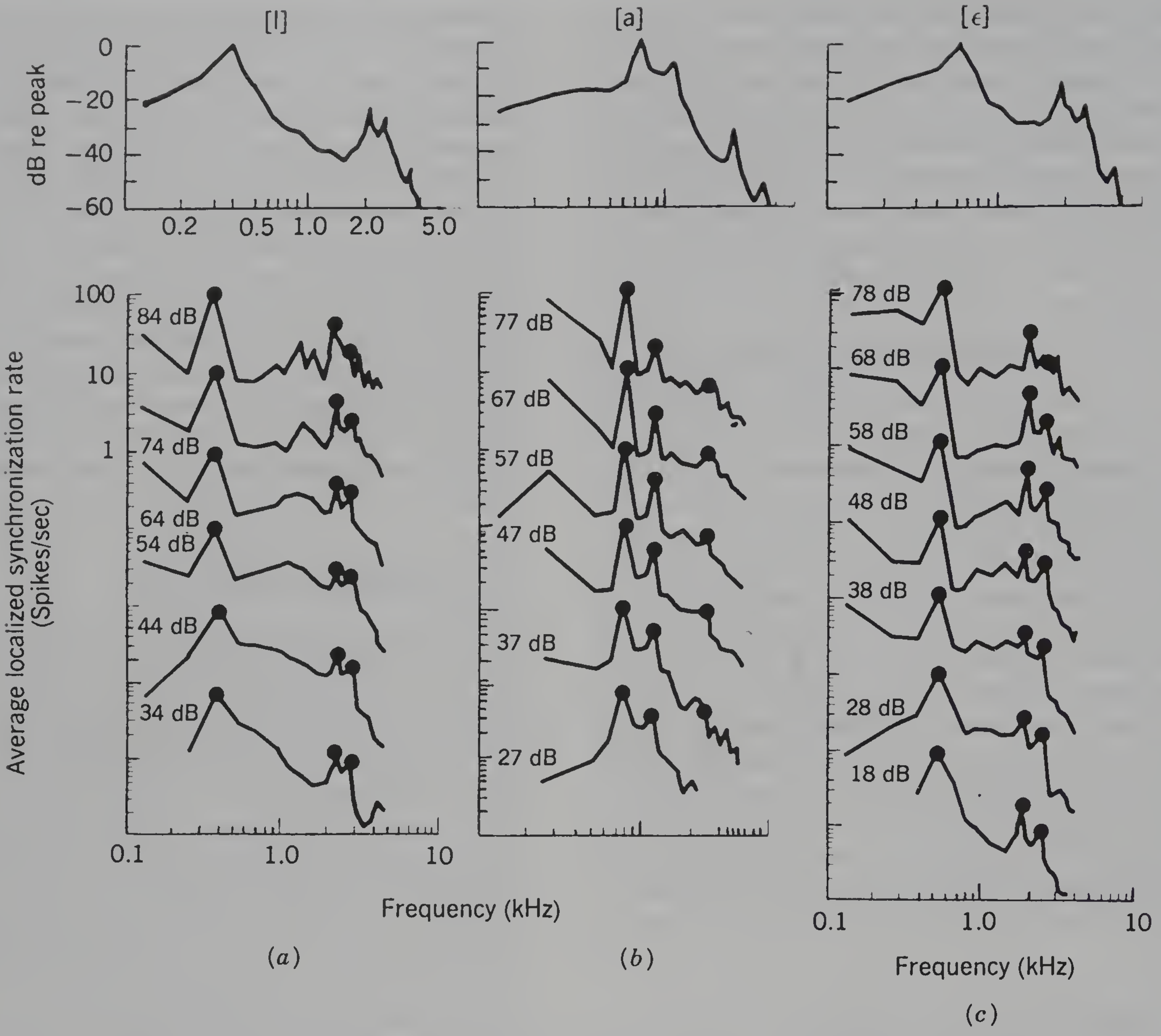


Fig. 3 Synchrony-place representation of vocalic spectra in the auditory nerve. Young and Sachs[13] have proposed the ALSR metric for computing the temporal representation of auditory nerve fibers to speech. The temporal activity of a nerve fiber goes through a matched filter corresponding to its characteristic frequency. Temporal activity in the range of the fiber characteristic frequency is passed through without attenuation, while temporal activity to remote frequencies is suppressed. Thus, the ALSR is a measure of the degree to which a fiber is synchronized to frequencies appropriate to its tonotopic affiliation. When computed in this fashion the temporal representation remains remarkably stable over a wide range of intensity levels. Illustrated are the ASLR representations across a 50–60-dB dynamic range for the vowels [I], [a], and [ϵ]. (Reprinted from Ref. 73, p. 509, with permission of MIT Press. Based on a figure from Ref. 13, with permission of the Acoustical Society of America.)

between the temporal activity and a neuron's filter characteristics, a requirement that currently lacks physiological support.

5.5 Onset Rate Information

The articulatory release associated with stop consonants [p, t, k, b, d, g] has a broadband spectrum and a rather abrupt onset, which evokes a marked flurry of activity across a wide characteristic frequency range of fibers. Another burst of activity occurs at the onset of voicing. Because the dynamic range of auditory-nerve fiber discharge is much larger during the initial rapid adaptation phase (0–10 ms),[75] there is relatively little or no saturation of discharge rate during this interval at high sound pressure levels.[76] In consequence, the onset spectra serving to distinguish the stop consonants[77] are adequately represented in the distribution of rate-place activity across the auditory nerve[67] over the narrow time window associated with articulatory release.

This form of rate information differs from the more traditional *average-rate* metric. The underlying parameter governing neural magnitude at onset is the probability of discharge over a very small time interval. This probability is usually converted into effective discharge rate. If the analysis window is sufficiently small (ca. 100 μs) the apparent rate can be exceedingly high (up to 10,000 spikes/s). Such high onset rates reflect two properties of the neural discharge—the high probability of firing correlated with stimulus onset and the small degree of variance for this first-spike latency. This measure of onset response magnitude is one form of *instantaneous* rate. Instantaneous, in this context, refers to the spike rate measured over an interval corresponding to the analysis bin width, which generally ranges between 10 and 1000 μs. This is in contrast to average rate, which reflects the magnitude of activity occurring over the entire stimulus duration. Average rate is essentially an integrative measure of activity, which counts spikes over relatively long periods of time, and weights each point in time equally. Instantaneous rate emphasizes the clustering of spikes over small time windows and is effectively a correlational measure of neural response. Activity that is highly correlated in time, upon repeated presentations, will over certain time intervals have very high instantaneous rates of discharge. Conversely, poorly correlated response patterns will show much lower peak instantaneous rates, whose magnitudes are close to that of the average rate. The distinction between integrative and correlational measures of neural activity is of critical importance for understanding how information in the auditory nerve is processed by neurons in the higher stations of the auditory pathway where information is passed on through the action of coincidence detection based on high degrees of temporal correlation.

5.6 Distributed Coding of the Spectrum in Auditory Nerve and Cochlear Nucleus

Synchrony-Distributed Information in Auditory Nerve The mechanism underlying synchrony suppression also ensures that temporal information pertaining to a speech signal's formant peaks (particularly F_1) is distributed across auditory-nerve fibers spanning a broad tonotopic range under many listening conditions. At low sound pressure levels formant-related synchrony is confined to fibers with characteristic frequencies close to the spectral peaks of the vowel [ɛ]. As the signal amplitude rises so does the number of fibers synchronized to F_1. At amplitude levels typical of conversational speech, most fibers with characteristic frequencies below 10 kHz are synchronized to the first formant.[13] Figure 4 shows the tonotopic distribution of synchrony in the auditory nerve to the vowel [ϵ]. There is a gap in the tonotopic distribution of F_1 synchrony, centered at 1800 Hz, the frequency of the second formant. This circumscribed population of fibers maintains its synchrony to F_2 by virtue of the relatively large amplitude of the second formant peak.

The broadening of auditory-nerve fiber frequency selectivity is the primary basis of this wide distribution of F_1 synchrony. At conversational levels, the filtering of auditory-nerve fibers becomes almost low pass, in contrast to the sharply bandpass characteristic observed in response to signals below 60 dB SPL. In consequence, low-frequency signals possess the capability of *capturing* the activity of auditory-nerve fibers most sensitive to much higher frequencies. The tonotopic distribution of synchrony to the lower formants of vocalic stimuli has been accurately simulated by Jenison and colleagues[78] using filter functions derived from the discharge-rate-based response areas (isointensity curves) of single auditory nerve fibers. In their simulation the filter functions change as a function of tonotopic frequency and sound pressure level. Only at intensities 60 dB or greater is there observed a spreading of F_1 synchrony, analogous to that reported by Young and Sachs.[13]

The broadening of auditory-nerve filtering is always toward the low-frequency portion of the spectrum, thus making low-frequency signals capable of *labeling* the neural response at any point along the tonotopic axis in terms of modulating (phase locking) the fiber discharge pattern. In the absence of this phase-locked pattern it would be difficult to identify the signal frequency driving a unit's response, as the auditory-nerve fiber discharge rate typically saturates 20–30 dB above rate threshold. Beyond this level there are few cues to distinguish excitation driven by frequencies close to the unit characteristic frequency from those more distant. The average-rate population profile in response to an intense low-frequency tone would, under these conditions, be virtually indistinguishable from the pattern evoked by a wideband noise.

In the mammalian auditory nerve there is a tendency for broad frequency selectivity to be associated with the presence of phase locking and for very sharp tuning to be correlated with discharge rate information. The tuning of auditory-nerve fibers is relatively broad (as measured by the unit's Q_{10} dB, defined as the fiber characteristic frequency divided by the bandwidth of the response function 10 dB above threshold) for low-characteristic-frequency (< 2 kHz) units (Q_{10} between 1 and 2) and rather sharp for fibers with characteristic frequencies greater than 4 kHz (Q_{10} between 4 and 20). Thus, at low intensities, where rate-place cues would be expected to play a dominant role in the coding of spectral information, the tuning of high-characteristic-frequency fibers is very

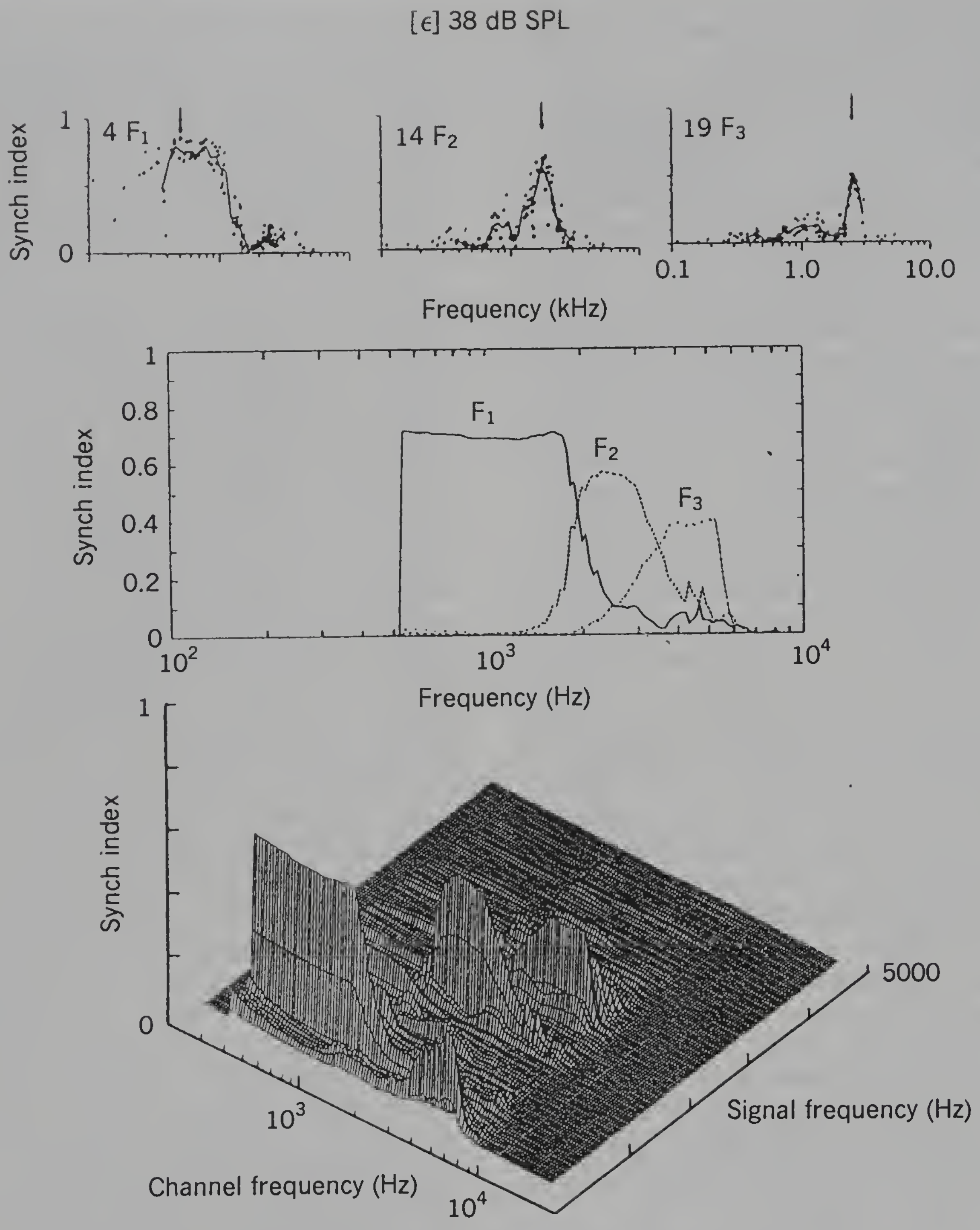


Fig. 4 Distributed representation of vocalic spectra in the auditory nerve. The tonotopic distribution of neural synchronization to the first three formants of the vowel [ɛ] at two different sound pressure levels. The top row shows the distribution of activity phase locked to the three lowest formants (F_1, F_2, and F_3) as recorded from the auditory nerve of the cat.[13] Notice that the distribution of activity to F_1 expands dramatically at 78 dB SPL, while islands of synchrony to F_2 and F_3 are preserved. The middle and lower rows show the output of a model based on physiological response areas. The distribution of synchrony corresponds very closely to that of the cat. (Reprinted from Ref. 78, pp. 780–781, with permission of the Acoustical Society of America.)

sharp. And this high degree of selectivity is maintained up to approximately 50–60 dB SPL. Such sharp tuning may be the natural consequence of a reliance on rate-place cues for the encoding of spectral information beyond the range of phase locking (i.e., above 4 kHz).

There are at least two principal consequences of distributing temporal information pertaining to the lower formants over a large number of auditory-nerve fibers. First, such distributed coding emphasizes spectral peaks at the expense of other components of the acoustic waveform. Lower amplitude components, particularly those located in the spectral trough between widely separated

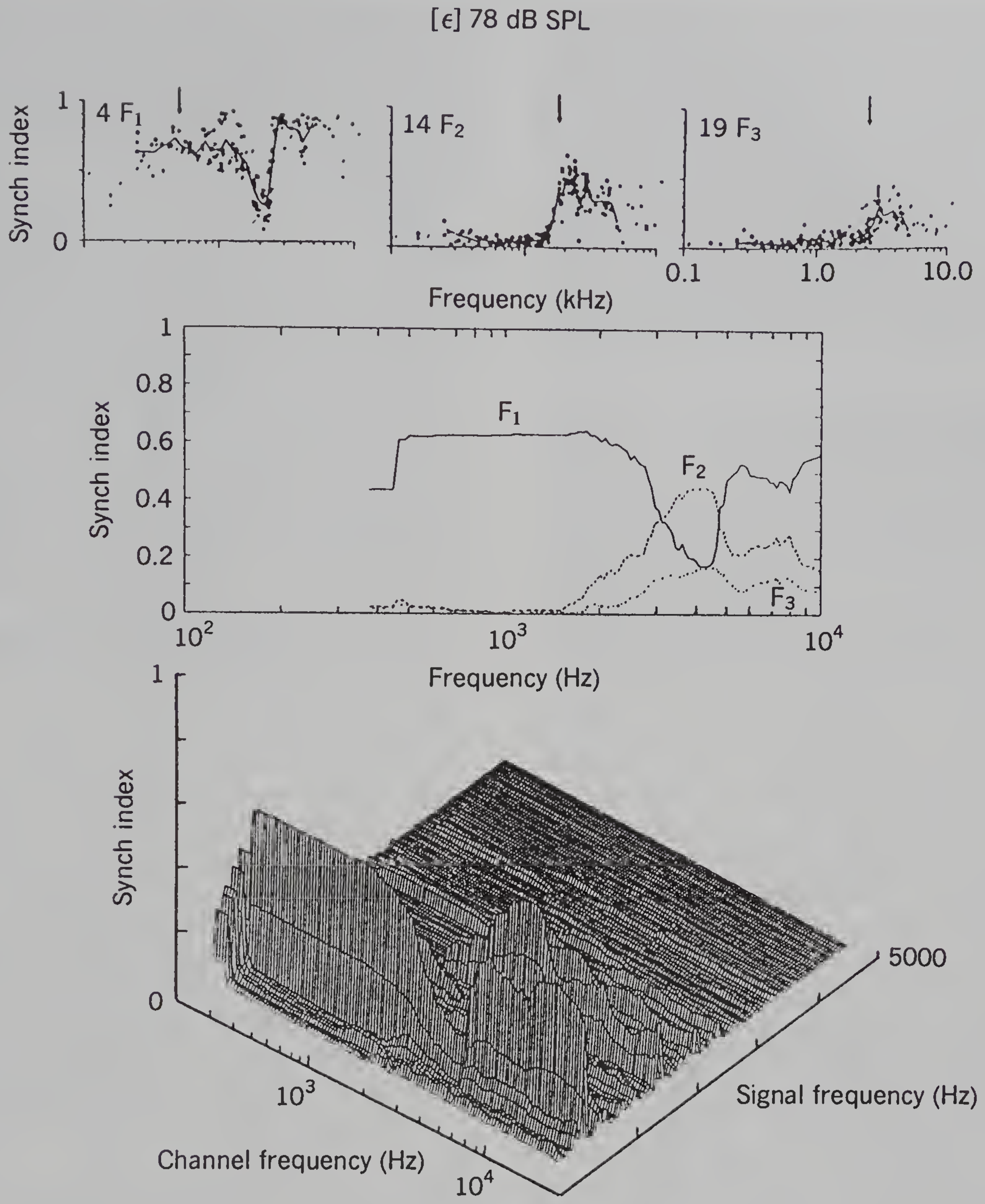


Fig. 4 (*Continued*)

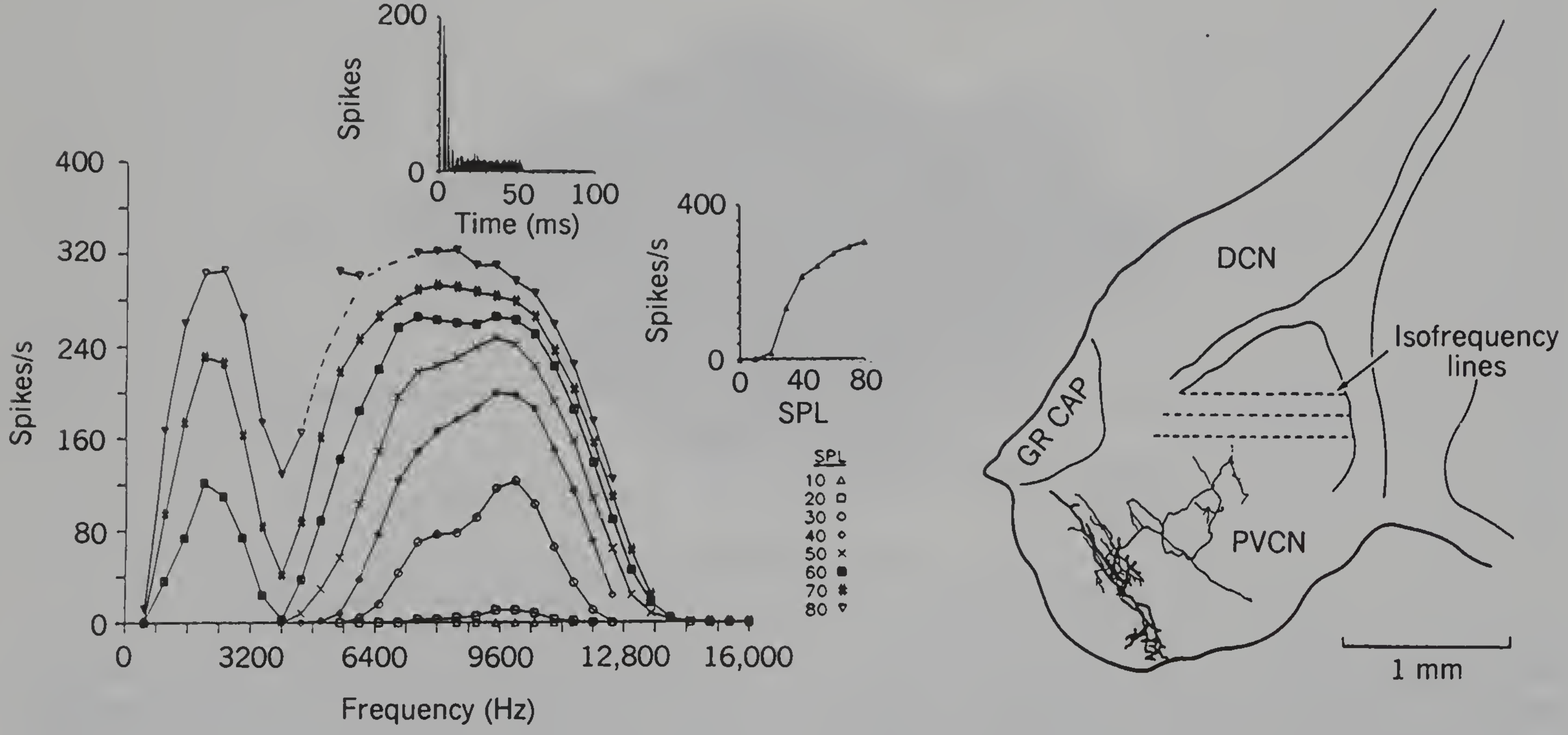
formant peaks (e.g., F_1 and F_2 for high and midfront vowels, F_2 and F_3 for back vowels) will not be represented clearly, if at all, in such neural activity patterns. In this sense, a distributed representation reduces the amount of data encoded in the peripheral neural activity pattern. It is similar to the *dominant frequency* model of Carlson et al.[79] (with extensions by Sinex and Geisler[80] and Ghitza[81]). This model assumes that only a small portion of the spectrum is behaviorally significant, and that the remainder is essentially "window dressing" with respect to biologically relevant information. This interpretation is consistent with the fact that listeners are capable of discriminating between synthetic vocalic stimuli of variable spectral complexity. Reduced spectra were hardly distinguishable from full-spectrum vowels, as long as the former contains the three lowest formants (each formant needs to be represented by at least three harmonics).[82] Hermansky[83] has shown that much of the experimental data pertaining to speech coding can be accommodated within a "perceptual linear predictive" framework in which the auditory system encodes only a highly smoothed, sparse representation of the signal spectrum.

A second consequence of distributed coding is the protection of the remaining information from the potentially deleterious effects of background noise and acoustic interference. Distribution of key, information-laden components (*info-elements*) of the acoustic signal across a broad tonotopic range reduces the probability that extraneous signals will severely compromise their representation. Since most interference is of a transient or narrow-band nature, the neural activity evoked by the info-elements will be preserved across much of the tonotopic axis, or will be interfered with only intermittently. It would be considerably more difficult to shield info-elements against sustained, broadband interference. Under these conditions, other strategies, based on common modulation patterns and binaural correlation may be required.[14,84,85]

Synchrony-Distributed Information in Posteroventral Cochlear Nucleus Essential to the dominant frequency model is the means by which common synchronous activity, distributed across the tonotopic axis, is recognized and integrated at the level of the cochlear nuclei and beyond. Most cells in the cochlear nuclei receive direct excitatory input from a relatively narrow tonotopic range of auditory-nerve fibers (an *isofrequency* projection). For this reason their ability to monitor the spread of common synchronous activity over a broad tonotopic range is limited. Such cells, however, may serve to transmit synchronous neural activity to higher centers possessing the required integrative machinery.

One class of neuron in the cochlear nucleus does appear to possess specific response properties required for the analysis of distributed temporal information. The onset choppers of the posteroventral division behave in a manner consistent with the dominant frequency model. In response to low-frequency sinusoidal or amplitude-modulated input, the average rate of these cells grows with intensity over a broad dynamic range, in tandem with the presumed tonotopic breadth of the active auditory-nerve projections. Thus, discharge rate in these units appears to reflect both the tonotopic extent of coherent temporal activity, as well as the intensity of the driving component, as shown in Fig. 5.

The onset chopper's ability to integrate coherent



(*a*) (*b*)

Fig. 5 Physiology and anatomy of a multipolar-stellate cell (onset chopper) in the posteroventral cochlear nucleus. The onset chopper units operate in many ways as coincidence detectors and intensity encoders. They have been anatomically associated with multipolar stellate cells of the posteroventral cochlear nucleus. (*a*) The response area, rate–intensity curve, and poststimulus–time histogram of the multipolarstellate (O_C) cell shown in (*b*). Characteristic frequency 10.4 kHz. Threshold 20 dB SPL. (*b*) A labeled O_C multipolar-stellate cell in the PVCN illustrating the extent of its dendritic field, and its orientation relative to the frequency plane of incoming auditory-nerve fibers. Horizontal lines indicate the approximate frequency isoclines for the auditory-nerve fiber projection. Low frequencies are ventral and high frequencies are dorsal (top). Dark lines indicate dendrites and light lines axonal projection. (Reprinted from Ref. 86, pp. 122–123, with permission of Springer Verlag.)

synchronous activity relies, in all probability, on two different properties. One concerns the broad frequency selectivity that reflects the wide tonotopic extent of the auditory nerve fiber projection. The other involves some form of *coincidence detection*. The latter refers to a situation in which a cell is sensitive to the relative arrival time of nerve impulses from two or more afferent inputs. Such neurons generally fire only if a certain minimum number of inputs discharge within a very small time window (on the order of tens to a few hundreds of microseconds). Onset choppers behave like coincidence detectors in that the variance of their initial-spike latency, as well as the variance of their phase-locked response to low-frequency modulations, is exceedingly small, far less variable than the analogous discharge measures for their auditory-nerve fiber projections. In this sense, onset units may act as correlators of afferent temporal activity, similar in function to that proposed for the cross-channel correlation model of Deng.[59]

Coincidence detection is a potentially very powerful means of protecting info-elements from the destructive effects of noise. Because noise implies the absence of sustained correlation, neural elements particularly responsive to correlated inputs are capable, in principle, of withstanding appreciable levels of acoustic interference, as long as the background is itself not highly correlated with the foreground signal. Under such circumstances the temporal pattern of neural activity will remain relatively unaffected at all but the lowest signal-to-noise ratios. Rhode and Greenberg[29] have examined the ability of various cell types in the cochlear nuclei to encode amplitude modulation in noise. Although all response classes exhibit impressive AM coding in noise, onset chopper units are among the most effective in withstanding the deleterious effects of acoustic interference.

In response to vocalic stimuli, synchrony to both F_1 and f_0 will be dispersed across a broad tonotopic range of auditory-nerve fiber inputs. At low SPLs the phase-locked response to f_0 will be more broadly distributed than F_1-synchronized activity. Fibers with characteristic frequencies below F_1, between F_1 and F_2, and above F_3 will exhibit appreciable phase locking to the fundamental frequency (f_0). This common synchrony to a low-modulation frequency could serve to encode important vocalic information in several different ways. In the presence of competing speech or other background sounds, the common f_0-synchronized activity serves as a means to identify tonotopically disjoint regions of the neural excitation pattern as originating from the same sound source. In noisy conditions the ability to track such common modulation patterns may act as a powerful cue for segregation and grouping of acoustic information into coherent *event streams*.[14] Moreover, the phase-locked manner of the *common source* coding would minimize the potentially disruptive effects of background noise. Even in the absence of acoustic interference, f_0 synchrony may serve to "bind" disparate portions of the tonotopic activity pattern together to form a single sound source.

At higher sound pressure levels the dispersion of f_0 synchrony decreases in the auditory nerve, while phase locking to F_1 expands appreciably. The tonotopic distribution of synchrony to f_0 is generally complementary to that of F_1. Rarely are fibers in a characteristic frequency region highly synchronized to both. Most fibers will be predominantly synchronized to either f_0 or F_1, with the exception of the relatively small proportion phase locked to F_2 and F_3. In consequence, the distribution of phase locking to f_0 can serve as a "negative" image of the tonotopic dispersion of F_1 synchrony, and vice versa. This complementary pattern has potentially significant consequences for coding vocalic stimuli in the upper reaches of the auditory pathway.

At the level of the cochlear nucleus virtually all cell types are capable of synchronizing to amplitude modulation within the frequency range of the speech fundamental (80–400 Hz). Onset chopper units exhibit the greatest precision of phase locking to low-frequency amplitude modulation, followed by pauser/buildup, chopper, onset locker, and primary-like units. Thus, low-frequency modulation is a potentially powerful cue for informational encoding, source segregation and signal extraction in noise in the cochlear nuclei.[87]

Coherent, low-frequency modulation of neural activity may also serve to provide a form of *pitch-synchronous* analysis. A major drawback of conventional Fourier analysis is its reliance on a fixed time window of arbitrary length. In the analysis of acoustic signals of variable or unknown periodicity such fixed-interval analyses may provide erroneous estimates of the spectrum. Synchronizing the analysis to the pitch period provides one means of circumventing this problem. It has been proposed that the first formant is estimated completely in the time domain via some form of auditory pitch synchronous analysis.[88] In this model each glottal pulse triggers the beginning of a new analysis frame, over which interval the number of major oscillations is counted. Vowel identification is based not on the absolute frequencies of $F_1, F_2, \ldots, F_n$, but rather on the number of modulations per pitch period. A related form of periodicity analysis has recently been proposed by Langner,[89,90] Patterson et al.,[62] and Holton et al.[36] In Langner's model, onset units act as pitch-synchronous triggers, and choppers serve as intrinsic oscillators upon which the analysis of the signal's temporal fine structure is ultimately made. Activity from these two neural populations is presumably correlated in a more central auditory region, such as the inferior colliculus.

Potential Significance of Distributed-Synchrony Representation for Speech Understanding It is a central paradox of the clinical literature that many individuals with a sensorineural hearing loss experience little or no difficulty understanding speech in relatively quiet, nonreverberant environments but suffer profound impairment in noisy or reverberant conditions. The impairment in sensitivity is generally localized to frequencies greater than 3 kHz, though the distribution of energy in the speech signal is biased toward frequencies *below* 3 kHz. Hearing aids, which essentially amplify the portion of the spectrum sustaining the greatest loss in sensitivity, rarely correct for the impairment under conditions where the problems are most acute—in noisy environments.

A possible explanation for this paradox is that the impairment in speech decoding is a consequence not of a loss of sensitivity, but rather is due to a reduction in the number of frequency channels over which correlated neural activity pertaining to specific speech features occur. In quiet conditions the absence of higher frequency channels would not be significant because most of the speech-relevant information would be reliably extracted from the low-frequency channels. But under noisy conditions the representation of acoustic features in these low-frequency channels may be sufficiently compromised as to require disambiguation in the higher frequency channels. A significant reduction in sensitivity in these channels would appreciably reduce the redundancy of neural information correlated with speech features and thereby diminish the integrity of the auditory representation associated with them. Consistent with this hypothesis is the fact that the psychoacoustic parameter most predictive of speech intelligibility in quiet is an individual's sensitivity to frequencies below 2 kHz, while in noise it is the sensitivity to signals above 2 kHz.[91]

6 CENTRAL AUDITORY MECHANISMS

At the higher stations of the auditory pathway, the ability of neurons to phase lock diminishes appreciably. In the inferior colliculus, units rarely synchronize to frequencies above 1 kHz,[92] and the upper limit of phase locking for units in the thalamic and cortical regions is about 200 Hz.[32]

The question naturally arises as to how information encoded in terms of synchrony in the auditory nerve and cochlear nucleus is converted into some form "understood" by the higher levels of the auditory pathway. What is the nature of this transformed temporal information, and what are the mechanisms responsible for the conversion?

In 1948, Lloyd Jeffress proposed a model for binaural analysis of sound based on coincidence detection and neural delay lines.[93] Jeffress's model was concerned exclusively with *time–intensity* trading for localization of sound sources. Licklider[94] extended this model into the domain of pitch analysis for complex sounds, using a mathematical technique known as *autocorrelation*. Autocorrelation is used to estimate the periodicity of an arbitrary waveform. In Licklider's model an exact copy of the original waveform (as filtered and neurally transduced in the auditory periphery) is "correlated" (multiplied) with the reference signal after a variable time delay. This delay is a consequence of neural circuitry that preserves the fine temporal details of the waveform but sends the signal down a longer path. This "high road" has, at various points along its route, connections (taps) to the more direct, "low road." Each successive tap represents an incremental delay. At each tap a correlational operation is performed. At most taps the original and delayed version will not be highly correlated since they are out of phase with one another. However, for periodic or quasi-periodic signals, such as voiced speech, there will be some delay interval over which the correlation between the two waveforms is high relative to that of other taps. This interval of maximum correlation provides an estimate of the signal periodicity. Because each tap has a well-defined location, the neural autocorrelation effectively converts a temporal measure of periodicity into a spatial representation, which in Licklider's model runs orthogonal to the primary, isofrequency representation of afferent activity and encodes the waveform periodicity pertaining to the signal's pitch and timbral properties. Licklider's approach has become increasingly popular as a means of summarizing the periodicity information contained with the auditory representation of complex signals.[56,60,62,95]

As elegant as the autocorrelational approach is for representing spectral and periodicity information there is, as yet, scant anatomical and physiological support for the requisite delay line. There is preliminary evidence in the barn owl for a delay line in nucleus laminaris, the avian homologue of the medial superior olive,[96] but without any direct support for neural correlation. There is some very preliminary evidence that the medial superior olive may contain a delay line useful for sound localization,[97] although it is unclear whether such a mechanism would contain delays sufficiently long (20–30 ms) required to pitch analysis.

Perhaps the most intriguing results come from a study of the inferior colliculus of the cat, in which a topographic representation of waveform periodicity was found, based on a rate measure of best modulation frequency.[98,99] This best modulation frequency representation spans a range between 50 and 500 Hz, and its spatial organization appears to cut across the tonotopic frequency axis, just as Licklider's model requires. However, the anatomical and physiological basis of this best modulation frequency map remains obscure. And it is not entirely clear whether such a coarse topographic representation would be adequate to account for the precision of pitch discrimination among human listeners (ca. 1 Hz at 1 kHz or 0.1%). Strong evidence has been adduced in the cat inferior colliculus for some form of correlational operation pertaining to binaural processing of low-frequency information,[92,100] and it is likely that such a mechanism also plays a role in the central analysis of pitch and formant information.

Binaural analysis also plays an important role in both localization of sound and noise reduction due to correlation of inputs from the two ears. A detailed discussion of sound localization can be found in Chapter 117.

Noise reduction through binaural correlation was originally proposed by Cherry[101] and has received extensive attention over the intervening years. Recently, Stern and Trahiotis[102] have developed a model of binaural correlation using a multiple delay line approach that is quite effective in reducing background noise. The physiological basis of such delay lines has yet to be explored.

Virtually any acoustic environment will create reflections of sound, which will reach the listener after a delay typically ranging between tens and hundreds of milliseconds. Listeners are rarely aware of this reverberation, because the reflections appear to be perceptually absorbed into the original signal. Although the acoustic conditions under which this *precedence effect* occurs has been extensively studied,[103–105] its physiological basis remains obscure. There is some evidence suggesting that neurons in the inferior colliculus are sensitive to reverberant signals, and have the capability of inhibiting their response under such conditions,[106] but the role of the more central pathway in handling reverberation is controversial.[107]

However, it is clear that whatever physiological mechanisms underlie the processing of reverberant signals, they play an important role in both speech communication and natural object identification. The spectrotemporal characteristics of a sound source are dramatically altered as a consequence of reverberation, yet listeners rarely encounter any difficulty understanding speech under such conditions. A fuller understanding of these mechanisms is essential for any general model of speech recognition and auditory perception.

Neurons in the thalamic radiations and auditory cortex of awake monkeys are capable of highly synchronizing to the fundamental frequency of synthetic speech sounds,[108] suggesting that the temporal capability of at least some central auditory neurons is sufficient to encode f_0. At this level of auditory processing such temporal information may be able to encode other properties of the speech signal, including the lower formants, duration and intensity.

In the upper reaches of the auditory pathway (the inferior colliculus, medial geniculate body, and auditory cortex), the discharge rate of neurons is appreciably lower than observed among cells in the cochlear nucleus and auditory nerve. This decrease in discharge rate may be at least partially ascribable to the anesthetized condition of the experimental animal. However, there is reason to believe that even in the awake, behaving animal, cortical and thalamic neurons are less excitable than units in the periphery. It has been suggested that this decrease in responsiveness reflects the prevalence of encoding information through coincidence detection, rather than via energy integration and that most of the spectral analysis and preliminary acoustic feature extraction is performed in the more caudal stations of the auditory pathway.[109] What remains to be done at higher levels is a synthesis of primitive elements into a unified, coherent picture of the acoustic stream. In essence, the role of the cortical regions of the auditory system may be to process auditory "events," whose features change much more slowly than the acoustic spectrum, thus alleviating the need to update the neural stream more frequently than several times a second.

This view is consistent with the relatively low spike rates of most neurons in the auditory cortex, which rarely fire more than 5–10 times per second. At such low rates, it is possible for cortical neurons to encode meaningful acoustic features and objects, and it may not be coincidental that the rate at which speech syllables and phonetic segments occur is within this range. Such low encoding rates would also allow integration with information derived from other sensory modalities and the motor system, which have similar time constants.

7 CONCLUSIONS

The traditional view of the ear as a linear signal analyzer is no longer tenable in light of recent research indicating substantial nonlinearities at all levels beyond the middle ear.[110–113] It is becoming increasingly clear that many of these nonlinearities are not merely artifacts of imperfect design or implementation but rather form an elegant system for processing acoustical information under real-world conditions, where extraction of communicationally relevant features in acoustically variable and uncertain conditions is commonplace.[114]

The focus of auditory theory is also changing, from a perspective oriented toward the coding of auditory-specific attributes (such as pitch), to a more dynamic, holistic one concerned with how sounds are processed under realistic conditions. The traditional categories of auditory experience—pitch, timbre, loudness, and spatial location—are now viewed as components of a larger gestalt in which the ear attempts to reconstruct the external world through interpretation of acoustic events.

There is also an increasing trend of integrating computational methods into auditory research, particularly with respect to modeling the anatomy and physiology of the auditory pathway. So far, most of the impetus for computational approaches has come from speech recognition research,[25,81,115] but such techniques will surely be used for clinical and basic research purposes in years to come.

REFERENCES

1. G. S. Ohm, "Über Die Definition Des Tones, Nebst Daran Geknupfter Theorie Der Sirene Und Ahnlicher Tonbildener Vorrichtungen," *Ann. Phys.*, Vol. 59, 1843, pp. 497–565.
2. H. L. F. von Helmholtz, *Die Lehre von Tonemfindungen als Physiologie Grundlage dur die Theorie der Musik*, F. Vieweg und Sohn, Braunschweige, 1863; translated as *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th ed., trans. by A. J. Ellis, Dover, New York, 1954 (reprint of 1897 edition).
3. R. Plomp, "The Ear as a Frequency Analyzer," *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 1628–1638.
4. N. Y. S. Kiang, T. Watenabe, E. C. Thomas, and L. F. Clark, *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*, MIT Press, Cambridge, MA, 1965.
5. I. C. Whitfield, *The Auditory Pathway*, William & Wilkins, Baltimore, 1967.
6. A. Seebeck, "Über Die Sirene," *Ann. Phys. Chem.*, Vol. 60, 1843, pp. 449–481.
7. W. Rutherford, "Tone Sensation with Reference to the Function of the Cochlea," *Lancet*, Vol. 76, 1898, pp. 389–394.
8. E. G. Wever, *Theory of Hearing*. Wiley, New York, 1949.
9. E. F. Evans, "The Dynamic Range Problem," in J. Syka and L. Aitkin (Eds.), *Neuronal Mechanisms in Hearing*, Plenum, New York, 1981, pp. 69–85.
10. W. S. Rhode and S. R. Greenberg, "Lateral Suppression and Inhibition in the Cochlear Nucleus of the Cat," *J. Neurophysiol.*, Vol. 71, 1994, pp. 493–519.
11. R. R. Pfeiffer and D. O. Kim, "Cochlear Nerve Fiber Responses: Distribution Along the Cochlear Partition," *J. Acoust. Soc. Am.*, Vol. 58, 1975, pp. 867–869.
12. M. B. Sachs and E. D. Young, "Encoding of Steady-State Vowels in the Auditory Nerve: Representation in Terms of Discharge Rate," *J. Acoust. Soc. Am.*, Vol. 66, 1979, pp. 470–479.
13. E. D. Young and M. B. Sachs, "Representation of Steady-State Vowels in the Temporal Aspects of the Discharge Patterns of Auditory-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 66, 1979, pp. 1381–1403.
14. A. S. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, MA, 1990.
15. S. Handel, *Listening*, MIT Press, Cambridge, 1989, Chapters 7 and 8.
16. S. McAdams, "Recognition of Sound Sources and Events," in S. McAdams and E. Bigand, (Eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition*, Oxford University Press, Oxford, 1993, pp. 146–198.
17. J. J. Gibson, *The Senses Considered as Perceptual Systems*, Houghton Miflin, Boston, 1966.
18. J. B. Nadol, "Comparative Anatomy of the Cochlea and Auditory Nerve in Mammals," *Hear. Res.*, Vol. 34, 1988, pp. 253–266.
19. D. K. Ryugo, "The Auditory Nerve: Peripheral Innervation, Cell Body Morphology and Central Projections," in D. B. Webster, A. N. Popper, and R. R. Fay (Eds.), *The Mammalian Auditory Pathway: Neuroanatomy*, Springer Verlag, New York, 1992, pp. 23–65.
20. S. Greenberg, "The Ear as a Speech Analyzer," *J. Phonet.*, Vol. 16, 1988, pp. 139–150.
21. J. C. Clarey, P. Barone, and T. J. Imig, "Physiology of Thalamus and Cortex," in A. N. Popper and R. R. Fay (Eds.), *The Mammalian Auditory Pathway: Neurophysiology*, Springer Verlag, New York, 1992, pp. 232–334.
22. J. Grey and J. Gordon, "Perceptual Effects of Spectral Modifications on Musical Timbres," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 1493–1500.
23. C. J. Darwin, "Perceptual Grouping of Speech Components Differing in Fundamental Frequency and Onset Time," *Q. J. Exper. Psych.*, Vol. 33A, 1981, pp. 185–207.
24. D. O. Kim, W. S. Rhode, and S. Greenberg, "Responses of Cochlear Nucleus Cells to Speech Signals: Neural Encoding of Pitch, Intensity and Other Parameters," in B. C. J. Moore and R. D. Patterson (Eds.), *Auditory Frequency Selectivity*, Plenum, New York, 1986, pp. 281–288.
25. T. Holton, "A Computational Approach to Recognition of Speech Features Using Models of Auditory Signal Processing," *Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 3,* Stockholm, Sweden, 1995, pp. 50–57.
26. J. F. Brugge, D. J. Anderson, J. E. Hind, and J. E. Rose, "Time Structure of Discharges in Single Auditory Nerve Fibers of the Squirrel Monkey in Response to Complex Periodic Sounds," *J. Neurophysiol.*, Vol. 32, 1969, pp. 386–401.
27. P. X. Joris and T. C. Yin, "Responses to Amplitude-Modulated Tones in the Auditory Nerve of the Cat," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 215–232.
28. R. Frisina, R. L. Smith, and S. Chamberlain, "Encoding of Amplitude Modulation in the Gerbil Cochlear Nucleus: I. Hierarchy of Enhancement," *Hear. Res.*, Vol. 44, 1990, pp. 99–122.

29. W. S. Rhode and S. R. Greenberg, "Encoding of Amplitude Modulation in the Cochlear Nucleus of the Cat," *J. Neurophysiol.*, Vol. 71, 1994, pp. 1797–1825.
30. R. Plomp, *Aspects of Tone Sensation*, Academic Press, London, 1976.
31. J. F. Schouten, "The Residue and the Mechanism of Hearing," *Proc. K. Ned. Akad. Wet.*, Vol. 43, 1940, pp. 991–999.
32. F. de Ribaupierre, M. H. Goldstein, Jr., and G. Yeni-Komshian, "Cortical Coding of Repetitive Acoustic Pulses," *Brain Res.*, Vol. 48, 1972, pp. 205–225.
33. C. E. Schreiner and J. V. Urbas, "Representation of Amplitude Modulation in the Auditory Cortex of the Cat. I. The Anterior Auditory Field (AAF)," *Hear. Res.*, Vol. 21, 1986, pp. 227–241.
34. E. R. Lewis, "Speculations about Noise and the Evolution of Vertebrate Hearing," *Hear. Res.*, Vol. 25, 1987, pp. 83–90.
35. A. W. Bronkhorst and R. Plomp, "Binaural Speech Intelligibility in Noise for Hearing-Impaired Listeners," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 1374–1383.
36. T. Holton, S. Love, and S. Gill, "Formant and Pitch Pulse Detection Using Models of Auditory Signal Processing," *Proceedings of the International Conference on Spoken Language Processing*, Banff, Canada, 1992, pp. 81–84.
37. M. C. Liberman, "Auditory-Nerve Response from Cats Raised in a Low-Noise Chamber," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 442–445.
38. C. D. Geisler, L. Deng, and S. Greenberg, "Thresholds for Primary Auditory Fibers Using Statistically Defined Criteria," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 1102–1109.
39. P. M. Sellick and I. J. Russell, "Intracellular Studies of Hair Cells: Filling the Gap Between Basilar Membrane Mechanics and Neural Excitation," in R. F. Naunton and C. Fernandez (Eds.), *Evoked Electrical Activity in the Auditory Nervous System*, Academic Press, New York, 1978, pp. 113–139.
40. S. Greenberg, "Acoustic Transduction in the Auditory Periphery," *J. Phonet.*, Vol. 16, 1988, pp. 3–17.
41. J. E. Rose, J. F. Brugge, D. J. Anderson and J. E. Hind, "Phase-Locked Response to Low-Frequency Tones in Single Auditory Nerve Fibers of the Squirrel Monkey," *J. Neurophysiol.*, Vol. 30, 1967, pp. 769–793.
42. T. B. Schalk and M. B. Sachs, "Nonlinearities in Auditory-Nerve Responses to Bandlimited Noise," *J. Acoust. Soc. Am.*, Vol. 67, 1980, pp. 903–913.
43. M. B. Sachs and N. Y. S. Kiang, "Two-Tone Inhibition in Auditory Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 1120–1128.
44. M. B. Sachs and E. D. Young, "Effects of Nonlinearities on Speech Encoding in the Auditory Nerve," *J. Acoust. Soc. Am.*, Vol. 68, 1980, pp. 858–875.
45. W. S. Rhode and P. H. Smith, "Characteristics of Tone-Pip Response Patterns in Relationship to Spontaneous Rate in Cat Auditory Nerve Fibers," *Hear. Res.*, Vol. 12, 1985, pp. 159–168.
46. B. H. A. Sokolowski, M. B. Sachs, and J. L. Goldstein, "Auditory Nerve Rate-Level Functions for Two-Tone Stimuli: Possible Relation to Basilar Membrane Nonlinearity," *Hear. Res.*, Vol. 41, 1989, pp. 115–124.
47. G. K. Yates, I. M. Winter, and D. Robertson, "Basilar Membrane Nonlinearity Determine Auditory Nerve Rate-Intensity Functions and Cochlear Dynamic Range," *Hear. Res.*, Vol. 45, 1990, pp. 203–220.
48. R. L. Smith, "Short-Term Adaptation in Single Auditory Nerve Fibers: Some Poststimulatory Effects," *J. Neurophysiol.*, Vol. 40, 1977, pp. 1098–1111.
49. E. M. Relkin and J. R. Doucet, "Recovery from Prior Stimulation. I: Relationship to Spontaneous Firing Rates of Primary Auditory Neurons," *Hear. Res.*, Vol. 55, 1988, pp. 215–222.
50. M. I. Miller and M. B. Sachs, "Representation of Voice Pitch in Discharge Patterns of Auditory-Nerve Fibers," *Hear. Res.*, Vol. 14, 1984, pp. 257–279.
51. S. Greenberg, "Possible Role of Low and Medium Spontaneous Rate Cochlear Nerve Fibers in the Encoding of Waveform Periodicity," in B. C. J. Moore and R. D. Patterson (Eds.), *Auditory Frequency Selectivity*, Plenum, New York, 1986, pp. 241–248.
52. L. Robles, M. A. Ruggero, and N. Rich, "Basilar Membrane Mechanics at the Base of the Chinchilla Cochlea. I. Input–Output Functions, Tuning Curves, and Response Phases," *J. Acoust. Soc. Am.*, Vol. 80, 1986, pp. 1364–1374.
53. R. F. Lyon, "Automatic Gain Control in Cochlear Mechanics," in C. D. Dallos, J. W. Geisler, C. Mathews, M. A. Ruggero, and C. S. Steele (Eds.), *The Mechanics and Biophysics of Hearing*, Springer, New York, 1990.
54. R. Meddis, "Simulation of Mechanical to Neural Transduction in the Auditory Receptor," *J. Acoust. Soc. Am.*, Vol. 79, 1986, pp. 702–711.
55. L. A. Westerman and R. L. Smith, "A Diffusion Model of the Transient Response of the Cochlear Inner Hair Cell Synapse," *J. Acoust. Soc. Am.*, Vol. 83, 1988, pp. 2266–2276.
56. R. F. Lyon, "Computational Models of Neural Auditory Processing," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1984, pp. 3611–3614.
57. B. Delgutte, "Analysis of French Stop Consonants Using a Model of the Peripheral Auditory System," in J. Perkell and D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes*, Lawrence Erlbaum, Hillsdale, NJ, 1987, pp. 163–177.
58. S. A. Shamma, "Speech Processing in the Auditory System. II: Lateral Inhibition and the Central Processing of Speech Evoked Activity in the Auditory Nerve," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 1622–1632.
59. L. Deng, C. D. Geisler, and S. Greenberg, "A Composite

Model of the Auditory Periphery for the Processing of Speech," *J. Phonet.*, Vol. 16, 1988, pp. 93–108.

60. R. Meddis and M. J. Hewett, "Virtual Pitch and Phase Sensitivity of a Computer Model of the Auditory Periphery: I. Pitch Identification," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 2866–2882.
61. R. Meddis and M. J. Hewett, "Virtual Pitch and Phase Sensitivity of a Computer Model of the Auditory Periphery: II. Phase Sensitivity," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 2883–2894.
62. R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, "Complex Sounds and Auditory Images," in Y. Cazal, L. Demany, and K. Horner (Eds.), *Auditory Physiology and Perception*, Pergamon, Oxford, 1992, pp. 429–446.
63. M. J. Hewitt, R. Meddis, and T. M. Shackleton, "A Computer Model of a Cochlear-Nucleus Stellate Cell: Responses to Amplitude-Modulated and Pure-Tone Stimuli," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 2096–2109.
64. M. J. Hewitt and R. Meddis, "A Computer Model of Amplitude-Modulation Sensitivity of Single Units in the Inferior Colliculus," *J. Acoust. Soc. Am.*, Vol. 95, 1994, pp. 2145–2159.
65. M. B. Sachs, R. L. Winslow, and B. H. Sokolowski, "A Computational Model for Rate-Level Functions from a Cat Auditory-Nerve Fibers," *Hear. Res.*, Vol. 41, 1989, pp. 61–69.
66. C. C. Blackburn and M. B. Sachs, "The Representation of the Steady-State Vowel Sound [ɛ] in the Discharge Patterns of Cat Anteroventral Cochlear Nucleus Neurons," *J. Neurophysiol.*, Vol. 63, 1990, pp. 1191–1212.
67. B. Delgutte and N. Y. S. Kiang, "Speech Coding in the Auditory Nerve: IV. Sounds with Consonant-Like Dynamic Characteristics," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 897–907.
68. W. H. Huggins, "A Phase Principle for Complex-Frequency Analysis and Its Implications in Auditory Theory," *J. Acoust. Soc. Am.*, Vol. 24, 1952, pp. 582–589.
69. D. J. Anderson, J. E. Rose, and J. F. Brugge, "Temporal Position and Discharges in Single Auditory Nerve Fibers within the Cycle of a Sine-Wave Stimulus: Frequency and Intensity Effects. *J. Acoust. Soc. Am.*, Vol. 49, 1971, pp. 1131–1139.
70. L. M. Kitzes, M. M. Gibson, J. E. Rose, and J. E. Hind, "Initial Discharge Latency and Threshold Considerations for Some Neurons in Cochlear Nucleus Complex of the Cat," *J. Neurophysiol.*, Vol. 41, 1978, pp. 1165–1182.
71. J. L. Goldstein and P. Srulovicz, "Auditory Nerve Spike Intervals as an Adequate Basis for Aural Spectrum Analysis," in E. F. Evans and J. P. Wilson (Eds.), *Psychophysics and Physiology of Hearing*, Academic Press, London, 1977, pp. 337–346.
72. P. Srulovicz and J. L. Goldstein, "A Central Spectrum Model: A Synthesis of Auditory-Nerve Timing and Place Cues in Monaural Communication of Frequency Spectrum," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 1266–1275.
73. S. Handel, *Listening*, MIT Press, Cambridge, 1989, p. 509.
74. H. Fletcher, *Speech and Hearing in Communication*, Van Nostrand, New York, 1953.
75. R. L. Smith and M. L. Brachman, "Operating Range and Maximum Response of Single Auditory Nerve Fibers," *Brain Res.*, Vol. 184, 1980, pp. 499–505.
76. M. B. Sachs, E. D. Young, and M. I. Miller, "Encoding of Speech Features in the Auditory Nerve," in R. Carlson and B. Granström, (Eds.), *Representation of Speech in the Peripheral Auditory System*, Elsevier, Amsterdam, 1982, pp. 115–130.
77. S. E. Blumstein and K. N. Stevens, "Perceptual Invariance and Onset Spectra for Stop Consonants in Different Vowel Environments," *J. Acoust. Soc. Am.*, 1980, Vol. 67, pp. 648–662.
78. R. Jenison, S. Greenberg, K. Kluender, and W. S. Rhode, "A Composite Model of the Auditory Periphery for the Processing of Speech Based on the Filter Response Functions of Single Auditory Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 773–786.
79. R. Carlson, G. Fant, and B. Granström, "Two Formant Modes, Pitch and Vowel Perception," in G. Fant and M. Tatham (Eds.), *Analysis and Perception of Speech*, Academic Press, London, 1975, pp. 55–82.
80. D. G. Sinex and C. D. Geisler, "Responses of Auditory-Nerve Fibers to Consonant-Vowel Syllables," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 602–615.
81. O. Ghitza, "Temporal Non-Place Information in the Auditory-Nerve Firing Patterns as a Front-End for Speech Recognition in a Noisy Environment," *J. Phonet.*, Vol. 16, 1988, pp. 109–124.
82. O. Kakusho, H. Hirato, K. Kato, and T. Kobayashi, "Some Experiments of Vowel Perception by Harmonic Synthesizer," *Acustica*, Vol. 24, 1971, pp. 179–190.
83. H. Hermansky, "Perceptual Linear Predictive Analysis for Speech," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 1738–1752.
84. F. A. Bilsen, "Pitch of Noise Signals: Evidence for a Central Spectrum," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 150–161.
85. J. W. Hall, J. H. Grose, and M. P. Haggard, "Comodulation Masking Release for Multicomponent Signals," *J. Acoust. Soc. of Am.*, Vol. 83, 1988, pp. 677–686.
86. W. S. Rhode and S. Greenberg, "Physiology of the Cochlear Nuclei," in A. N. Popper and R. R. Fay, (Eds.), *The Mammalian Auditory Pathway: Neurophysiology*, Springer Verlag, New York, 1992, pp. 94–152.
87. A. R. Palmer, "The Representation of the Spectra and Fundamental Frequencies of Steady-State Single- and Double-Vowel Sounds in the Temporal Discharge Patterns of Guinea-Pig Cochlear-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 1412–1426.
88. B. Scott, "Temporal Factors in Vowel Perception," *J. Acoust. Soc. Am.*, Vol. 60, 1976, pp. 1354–1360.

89. G. Langner, "Physiological Properties of Units in the Cochlear Nucleus Are Adequate for a Model of Periodicity Analysis in the Auditory Midbrain," in J. Syka and R. B. Masterton, *Auditory Pathway: Structure and Function*, Plenum, New York, 1988, pp. 207–212.
90. G. Langner, "Periodicity Coding in the Auditory System," *Hear. Res.*, Vol. 60, 1992, pp. 115–142.
91. G. F. Smoorenburg, "On the Limited Transfer of Information with Noise-Induced Hearing Loss," *Acta Otolaryngol.*, Vol. 469, 1990, pp. 38–46.
92. T. C. T. Yin and S. Kuwada, "Neuronal Mechanisms of Binaural Interaction," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Dynamic Aspects of Neocortical Function*, Wiley, New York, 1984, pp. 263–313.
93. L. A. Jeffress, "A Place Theory of Sound Localization," *J. Comp. Physiol. Psych.*, Vol. 41, 1948, pp. 35–39.
94. J. C. R. Licklider, "A Duplex Theory of Pitch Perception," *Experientia*, Vol. 7, 1951, pp. 129–133.
95. M. Slaney and R. Lyon, "On the Importance of Time—A Temporal Representation of Sound," in M. Cooke, S. Beet, and M. Crawford (Eds.), *Visual Representations of Speech Signals*. Wiley, Chichester, 1993, pp. 95–116.
96. C. E. Carr and M. Konishi, "Axonal Delay Lines Create Maps of Interaural Phase Differences in the Owl's Brainstem," *Proceedings of the National Academy of Science*, Vol. 85, 1987, pp. 8311–8315.
97. P. H. Smith, P. X. Joris, and T. C. Yin, "Projections of Physiologically Characterized Spherical Bushy Cell Axons from the Cochlear Nucleus of the Cat: Evidence for Delay Lines to the Medial Superior Olive," *J. Comp. Neurol.*, Vol. 331, 1993, pp. 245–260.
98. G. Langner and C. Schreiner, "Periodicity Coding in the Inferior Colliculus of the Cat. I. Neuronal Mechanisms," *J. Neurophysiol.*, Vol. 60, 1988, pp. 1799–1822.
99. C. Schreiner and G. Langner, "Periodicity Coding in the Inferior Colliculus of the Cat. II Topographical Organization," *J. Neurophysiol.*, Vol. 60, 1988, pp. 1823–1840.
100. T. C. Yin, J. C. Chan, and L. H. Carney, "Effects of Interaural Time Delays of Noise Stimuli on Low-Frequency Cells in the Cat's Inferior Colliculus. III. Evidence for Cross-Correlation," *J. Neurophysiol.*, Vol. 58, 1987, pp. 562–583.
101. C. Cherry, *On Human Communication*, Wiley, New York, 1957.
102. R. Stern and C. Trahiotis, "Models of Binaural Interaction," in B. C. J. Moore (Ed.), *Handbook of Perception and Cognition, Volume 6: Hearing*, Academic Press, San Diego, 1996, pp. 347–387.
103. H. Wallach, E. B. Newman, and M. R. Rosenzweig, "The Precedence Effect in Sound Localization," *Am. J. Psych.*, Vol. 62, 1949, pp. 315–336.
104. M. B. Gardner, "Historical Background of the Haas and/or Precedence Effect," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 1243–1248.
105. J. Blauert, G. Canevet, and T. Voinier, "The Precedence Effect: No Evidence for an 'Active' Release Process Found," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 2581–2586.
106. T. C. Yin, "Physiological Correlates of the Precedence Effect and Summing Localization in the Inferior Colliculus of the Cat," *J. Neurosci.*, Vol. 14, 1994, pp. 5170–5186.
107. J. L. Cranford and M. Oberholtzer, "Role of Neocortex in Binaural Hearing in the Cat: II. The Precedence Effect in Sound Localization," *Brain Res.*, Vol. 111, 1974, pp. 225–239.
108. M. Steinschneider, J. Arezzo, and H. Vaughan, "Speech Evoked Activity in the Auditory Radiations and the Cortex of the Awake Monkey," *Brain Res.*, Vol. 252, 1982, pp. 353–365.
109. M. Abeles, "Role of the Cortical Neuron: Integrator or Coincidence Detector? *Israel J. Med. Sci.*, Vol. 18, 1982, pp. 83–92.
110. N. Y. S. Kiang, "Peripheral Neural Processing of Auditory Information," in J. M. Brookhart and V. Mountcastle (Eds.), *Handbook of Physiology—The Nervous System, Vol. III, Sensory Processes, Part 2*, American Physiological Society, Bethesda, MD, 1984, pp. 639–674.
111. J. O. Pickles, *An Introduction to the Physiology of Hearing*, 2nd ed., Academic Press, London, 1988.
112. M. Ruggero, "Physiology and Coding of Sound in the Auditory Nerve," in A. N. Popper and R. R. Fay (Eds.), *The Mammalian Auditory Pathway: Neurophysiology*, Springer Verlag, New York, 1992, pp. 34–93.
113. D. Irvine, "Physiology of the Auditory Brainstem," in A. N. Popper and R. R. Fay (Eds.), *The Mammalian Auditory Pathway: Neurophysiology*, Springer Verlag, New York, 1992, pp. 153–231.
114. C. Brown, "The Perception of Vocal Signals by Blue Monkeys and Grey-Cheeked Mangabeys," *Experiment. Biol.*, Vol. 45, 1986, pp. 145–165.
115. N. Morgan, H. Bourlard, S. Greenberg, and H. Hermansky, "Stochastic Perceptual Auditory-Event-Based Models for Speech Recognition," *Proceedings of the International Congress on Spoken Language Processing*, Yokohama, Japan, 1994, pp. 1943–1946.

105

ACOUSTICAL CHARACTERISTICS OF THE OUTER EAR

EDGAR A. G. SHAW

1 INTRODUCTION

The outer ears and the head are components of a complex acoustical antenna system that couples the eardrums to the external sound field. This system is best characterized in terms of the free-field response measured at the eardrum and the interaural differences in response, as functions of source direction, that provide the physical basis for sound localization. The outer ear protects the eardrum from mechanical damage but also enhances the coupling between the eardrum and the sound field and contributes substantially to the directionality of the system, especially at high frequencies where the normal modes of the concha come into play. The acoustical functions of the various components of the outer ear have been elucidated through the development of physical models with characteristics designed to match those of the average human ear. When sounds are presented to the ear through earphones, an artificial system is created with complex acoustical characteristics that point to the need for real ear calibrations based on sound pressure measurements in the ear canal.

2 ANATOMICAL FEATURES

A descriptive diagram of the human outer ear is presented in Fig. 1 that also shows the principal positions at which sound pressure measurements have been made. In real ears, the canal is an irregular tubular structure with a sinuous central axis and a cross-sectional area that is sharply tapered in the region approaching and adjoining the eardrum, which, in fact, defines the shape of the canal at the inner end.[1] For this reason, and since there is no clear boundary between the concha and the canal, it is difficult to assign a precise meaning to canal length. However, at frequencies below 8 kHz, the primary acoustical characteristics of the average canal are well represented by a simple cylindrical cavity 7.5 mm in diameter and 22.5 mm in length (volume 1 cm^3) terminated by a pistonlike eardrum that is perpendicular to the axis.[2]

The concha, with a volume of approximately 4 cm^3, is a broad shallow cavity that is tightly coupled to the canal. This cavity is partially divided by the crus helias and the upper part, the cymba, is connected to the fossa. These features clearly have specific acoustical attributes[3] whereas the structures extending from the concha, such as the helix, the antihelix, and the lobule, seem to function collectively as a simple flange.[4] In this respect, the human ear is markedly different from many other mammalian ears in which the pinna extension is a significant conical structure under muscular control. Other differences are indicated in Table 1, which brings together some average dimensions for human, cat, and guinea pig ears. As can be seen, the concha volume is relatively large in the human ear. The outer ear is very shallow in infant ears but attains adult characteristics within 2 years.[5]

The pinna is a flexible cartilaginous structure that can generally be treated as a hard boundary. However, when the human pinna is brought into contact with a supra-aural earphone, the finite compliance of the concha wall is evident in an apparent increase in enclosed volume at frequencies below 300 Hz.[2] Similarly, the sound attenuation attainable with earplugs may be reduced when the plug is confined to the outer part of the canal where the skin is supported by cartilage rather than bone.

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

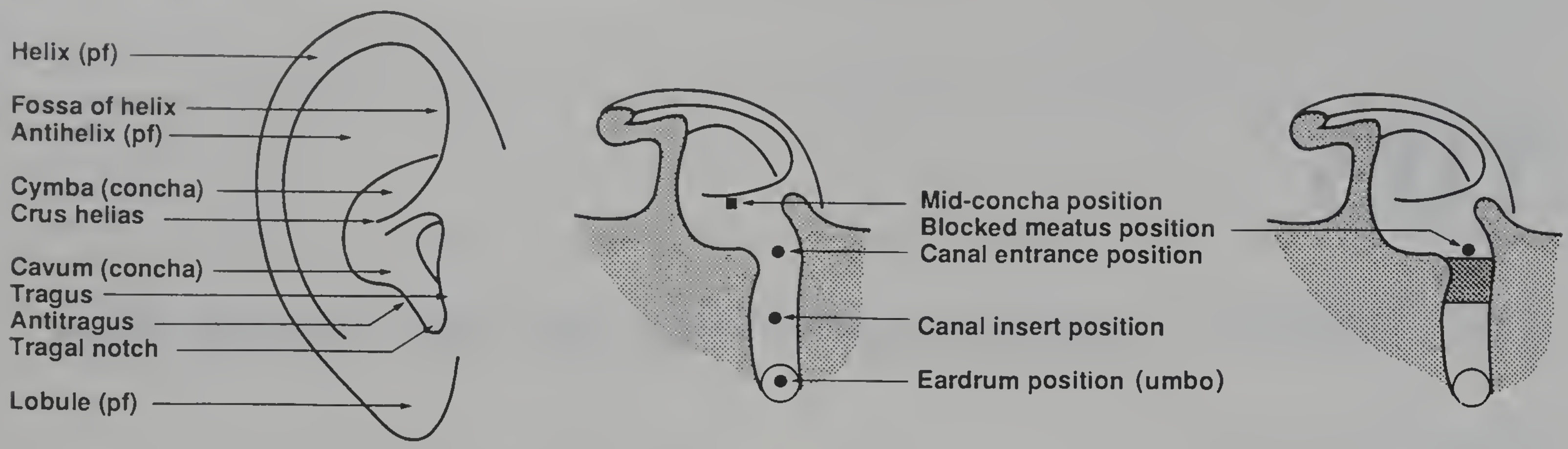


Fig. 1 Full view of human outer ear and horizontal cross sections showing five measurement positions.

3 TRANSFORMATION OF SOUND PRESSURE FROM FREE FIELD TO EARDRUM

The outer ear, head, neck, and torso are components of a complex acoustical antenna system that couples the middle and inner ear to the external sound field. This system can best be characterized in terms of the free-field response measured at the eardrum.

3.1 Transformation in the Horizontal Plane

Figure 2 shows average values of sound pressure level at the human eardrum, with respect to the free-field level at the center-head position, for plane waves approaching the head in the horizontal plane. This self-consistent family of response curves is based on a synthesis of experimental data from many laboratories[6,7] and is supported by other data.[8,9] Common to all the curves is the peak at 2.6 kHz, the primary resonance frequency of the average outer ear. Here the pressure gain attains its largest values. On the high-frequency side of the peak, resonance in the concha sustains the response for almost one octave. The steady decline between 5 and 10 kHz is peculiar to the horizontal plane and is associated with the mode structure of the concha.[3]

In the anterior sector, the directionality displayed in Fig. 2 is in broad agreement with calculations based on elementary diffraction theory where the outer ear is treated as a simple pressure detector at the surface of a hard spherical head.[2,6] At other angles of incidence, the agreement is at best approximate, especially at high frequencies. The difference in response between 45° and 135° azimuth, amounting to 11 dB at 4.5 kHz, is particularly significant. This lack of front–back symmetry, which provides a physical basis for localization in the lateral sector, can be attributed to diffraction by the pinna extension.[4]

The pure-tone transformation curves for individual subjects can be expected to be distributed about the average values with standard deviations of 1 dB or less below 500 Hz rising to 5 dB or more above 5 kHz. Particularly large variations are to be expected at high frequencies and in the shadow zone. Substantial smoothing occurs where pure tones are replaced by third-octave bands of noise.

3.2 Directionality in the Human Ear

Families of transformation curves comparable to those shown in Fig. 2 are not available for angles of incidence above and below the horizontal plane. It is, however, clear that variations in high-frequency response with

TABLE 1 **Average Outer Ear Dimensions**

	Canal Length (cm)	Canal Volume (cm^3)	Concha Depth (cm)	Concha Volume (cm^3)	Pinna Length (cm)	Interaural Distance (cm)
Man	2.25	1.0	1.0	3.8	6.7	17.5
Cat	1.5	0.23	1.0	0.3	4	10
Guinea pig	0.9	0.045	0.28	0.055	2.5	4

Source: After Ref. 2.

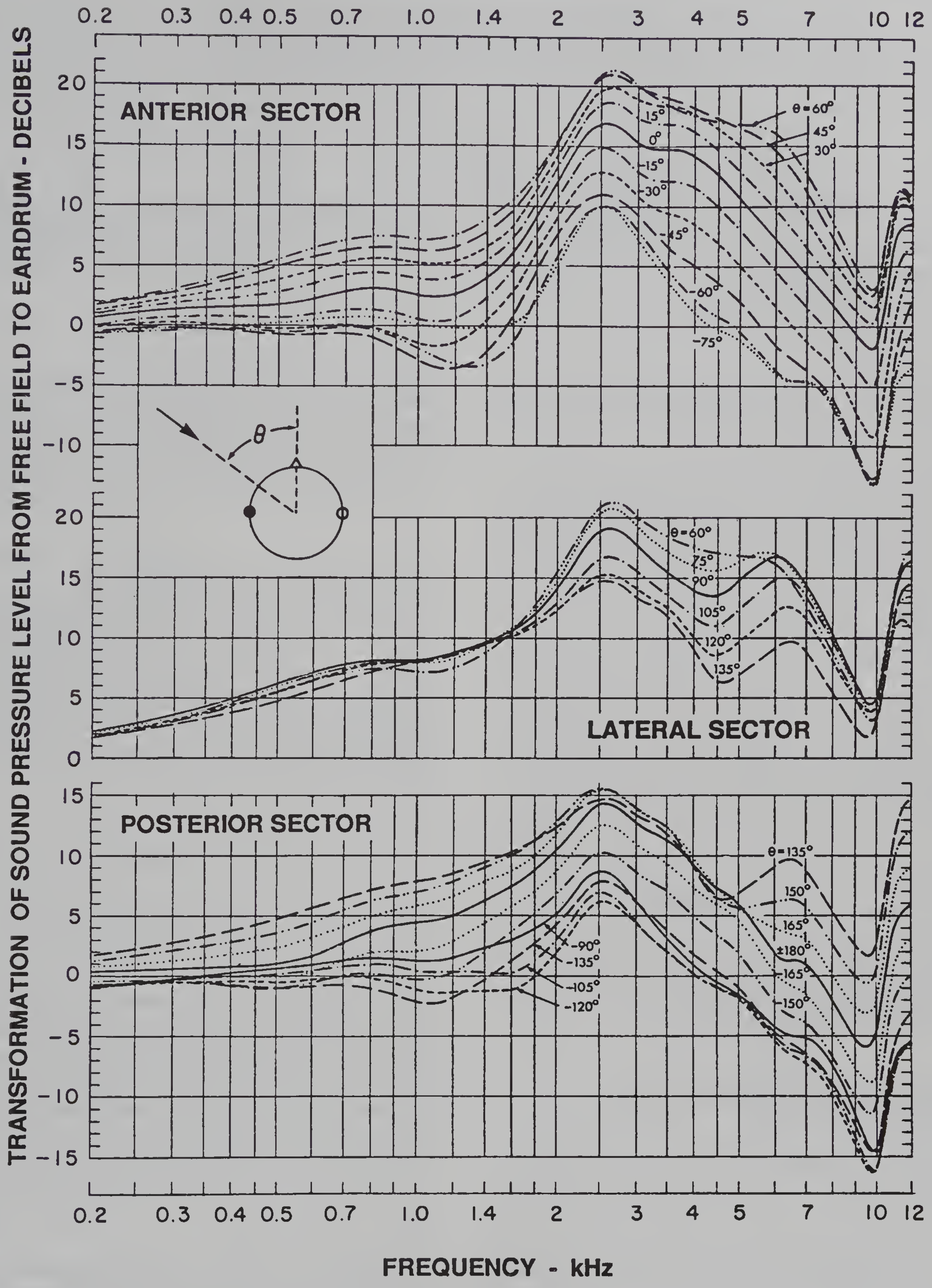


Fig. 2 Transformation of sound pressure level from the free field to the human eardrum in the horizontal plane as a function of frequency at 15° intervals of azimuth θ. Curves show average values based on data from 12 studies. (After Ref. 6).

TABLE 2 Directions of High Sensitivity in Human Ear: Approximate Angular Positions of Zone Centers and Angular Magnitudes of Zones

Frequency (kHz)	4	6	8^a	8^b	10	12^a	12^b	14
Azimuth, deg	40	60	80	30	70	10	90	15
Elevation, deg	10	20	30	−40	−20	−10	0	10
Magnitude, deg	±40 azimuth, ±60 elevation	±40	±30	±20	±25	±20	±15	±20

[a] Primary zone.
[b] Secondary zone.
Source: Based on Ref. 10.

source elevation[3,8] provide substantial spectral cues that are available for localization, especially in the median plane.

Some major high-frequency characteristics have recently been brought into focus in terms of contours of equal sensitivity presented on a spherical surface.[10] Similar patterns of directionality were found with all six subjects studied though the frequency associated with any particular pattern varied appreciably from subject to subject. Table 2 shows the approximate average directions and sizes of the primary and secondary zones of high sensitivity found in this study, where the levels at the zone boundaries are 5 dB below the highest sound pressure levels measured at each frequency. At 4 kHz, high sensitivity is maintained from 0° to 80° azimuth and from −40° to +80° in elevation. At 6 kHz, there is a decrease in zone size and movement of the zone center to greater angles of azimuth and elevation. At 8 kHz, the pattern becomes bipolar with zones of high sensitivity above and below the horizontal plane. The pattern is also bipolar at 12 kHz with the zones of high sensitivity now widely separated in azimuth and centered near the horizontal plane. These features provide substantial spectral cues that are available for sound localization.

3.3 Sound Pressure Transformation in Various Mammalian Ears

In many respects, the transformation characteristics measured in cats, guinea pigs, chinchillas, and other mammals are similar to those found in humans taking into account the differences in pinna dimensions.[2] At high frequencies, however, the patterns of directionality associated with the conical pinna extension are notably different from those of the human ear. For example, in cats with pinnas in the "alert" position, the direction of maximum sensitivity generally lies close to the frontal axis (e.g., azimuth ~ 30°, elevation ~ 20°), varying only modestly with frequency, and the sensitive zone systematically decreases in size between 4 and 20 kHz consistent with reception through a simple aperture of appropriate diameter (6 cm).[11] Nevertheless, refined measurements of sound pressure transformation from the free field to the cat eardrum display fine structure in the response curves not unlike that observed with human subjects.[12]

3.4 Interaural Time Differences

In calculations of interaural phase differences (IPDs) and interaural time differences (ITDs) based on diffraction theory, it is customary to treat the human head as a rigid sphere and the ears as simple pressure detectors at opposite ends of a diameter representing the interaural axis.[13] This treatment leads to the following asymptotic values of ITD at low and high frequencies:

$$\tau_{\mathrm{lf}} = \frac{a}{c}\,(3 \sin \theta), \tag{1}$$

$$\tau_{\mathrm{hf}} = \frac{a}{c}\,(\sin \theta + \theta), \tag{2}$$

where a is the radius of the sphere (~87.5 mm), c is the velocity of sound in air, and θ is the azimuth of the incident waves. The high-frequency asymptote agrees with the frequency-independent value of ITD stemming from geometrical acoustics where sound travels in a straight line to the left ear but bends to follow the spherical head as it approaches the right ear. In the transition region (0.5–2 kHz), it is necessary to distinguish between phase delay and group delay since the system is dispersive. (See also Chapter 117.)

Measurements of IPD with low-frequency tones and measurements of ITD with broadband high-frequency signals, such as clicks, have provided strong support for Eqs. (1) and (2).[2] Nevertheless, the actual high-frequency ITD for individual subjects must surely contain substantial frequency-dependent and direction-dependent variations in response corresponding to the sharp peaks and valleys seen in the amplitude domain.[14,15] Recent measurements of interaural envelope delays with signals of moderate bandwidth (0.22 octave) and center frequencies between 4 and 12 kHz provide mean values in broad agreement with Eq. (2) with individual values clustering about the mean with standard deviations of the order of 30%.[15]

4 ACOUSTICAL PARAMETERS, WAVE CHARACTERISTICS, AND PHYSICAL MODELS OF THE HUMAN OUTER EAR

Sound pressure measurements in real ears (Refs. 3, 6, 8, 10, 16, 17), experiments with replicas and physical models,[2–4] impedance measurements,[2,3] mathematical modeling,[1,18] and instrument technology and the quest for artificial ears[2,3,9] have all played a part in the formation of a comprehensive acoustical description of the human outer ear.

4.1 Eardrum Impedance

The impedance presented to the ear canal by the eardrum strongly affects the free-field response of the ear in the vicinity of resonance and is crucial when the ear is driven by an earphone inserted into the canal.

The parameter values presented in Table 3 are updated estimates of average values based on data from more than 20 studies.[2,3,16,19] Below 2 kHz, the original data are the mean and median values of R_d and X_d obtained in direct measurements of eardrum impedance. From 3 to 10 kHz, the primary data are the mean and median values of the standing-wave ratio (SWR) obtained from measurements of sound pressure distribution in the ear canal. The values of SWR presented in Table 3 indicate the general trend of these data. The corresponding values of R_d and X_d are calculated values based on the simplifying assumptions that the human ear canal can be treated as a uniform cylinder with a diameter of 7.5 mm (characteristic acoustic impedance $R_0 = 94$ g/cm$^4 \cdot$ s), and that R_d is essentially constant between 3 and 10 kHz. In the transition region (2–3 kHz), the values of R_d and X_d are consistent with the free-field response of the human ear at resonance.

At frequencies greater than 1 kHz, the eardrum impedance curves for individual subjects can be expected to display appreciable complexity reflecting the acoustical characteristics of the middle ear cavities and the dynamics of the eardrum, ossicular chain, and cochlea.

4.2 Normal Modes

The normal modes of the human outer ear provide the key to its high-frequency characteristics especially its directionality. It is, in fact, sufficient to study the modes of the concha and pinna extension in isolation since the directionality measured at the center of a rigid airtight plug closing the entrance to the ear canal (Fig. 1, blocked-meatus position) is essentially the same as that measured at the eardrum.[17]

Figure 3 shows the average characteristics of five modes based on data from 10 subjects measured under such conditions.[3] Mode 1, at 4.3 kHz, is a simple depth resonance with uniform pressure across the base of the concha.[4] All other modes have pressure distributions that are essentially transverse with nodal surfaces separating zones that are in phase opposition. At grazing incidence, mode 2, at 7.1 kHz, and mode 3, at 9.6 kHz, are best excited at approximately 75° elevation. In contrast, mode 4, at 12.1 kHz, and mode 5, at 14.4 kHz, are strongly excited from the front. In effect, the ear behaves as a vertical dipole under the control of modes 2 and 3 and as a horizontal dipole under the control of modes 4 and 5. These characteristics are consistent with the free-field directionality patterns measured in human subjects[10] (see Table 2).

4.3 Physical Models

The data presented in Fig. 3 pave the way for the construction of physical models with relatively simple geometry that have approximately the same mode frequencies, pressure distributions, directionality, and excitation as the average human ear.[3]

The essential characteristics of the depth resonance of the concha (mode 1 in Fig. 3) can be matched with a cylindrical cavity 22 mm in diameter and 10 mm in depth inclined at 30° to a rigid plane representing the

TABLE 3 Estimated Average Values of Ear Canal Standing-Wave Ratio (SWR), Eardrum Resistance (R_d), and Eardrum Reactance (X_d) for Normal Human Ears at Various Frequencies

Parameter	0.1 kHz	0.2 kHz	0.3 kHz	0.5 kHz	0.7 kHz	1 kHz	2 kHz	3 kHz	5 kHz	7 kHz	10 kHz
SWR, dB								13	14.5	19	21
R_d, g/cm$^4 \cdot$ s	490	430	390	350	330	320	390	420	400	400	400
X_d, g/cm$^4 \cdot$ s	−2800	−1400	−950	−550	−360	−170	−80	−30	200	410	510

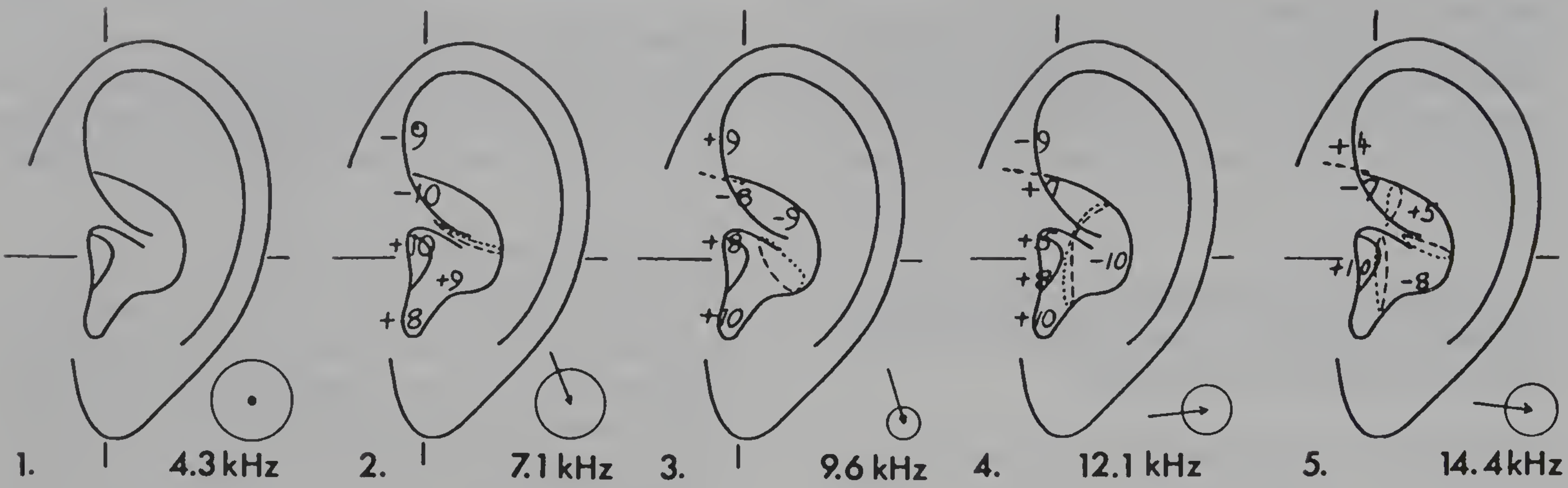


Fig. 3 Average transverse pressure distributions and resonance frequencies of five normal modes in human ears with ear canal closed at entrance showing nodal surfaces (broken lines) and relative pressure (numerals). Circles indicate relative degrees of excitation and arrows show directions of maximum response at grazing incidence. (After Ref. 3.)

surface of the head surrounding the ear (the *circumaural plane*). The cavity opening is partially surrounded by a simple flange, representing the pinna extension, which improves the coupling between the cavity and the sound field while suppressing reception from the rear.[4] To proceed beyond 5 kHz, greater geometrical complexity is required.[3] As a step in this direction, it is convenient to replace the cylindrical cavity with one that is rectangular in form. The first horizontal resonance (mode 4) is then readily tuned by adjusting the breadth of the cavity. To tune mode 2, with its vertical pressure distribution and remarkably low resonance frequency, it is essential to introduce a partial barrier representing the crus helias. This substantially increases the effective length of the cavity without increasing its volume. The presence of mode 3 is dependent on the introduction of a small auxiliary channel, connected to the main cavity, to represent the fossa. The final shape of the concha–fossa cavity is shown inset in panel (*a*) of Fig. 4.

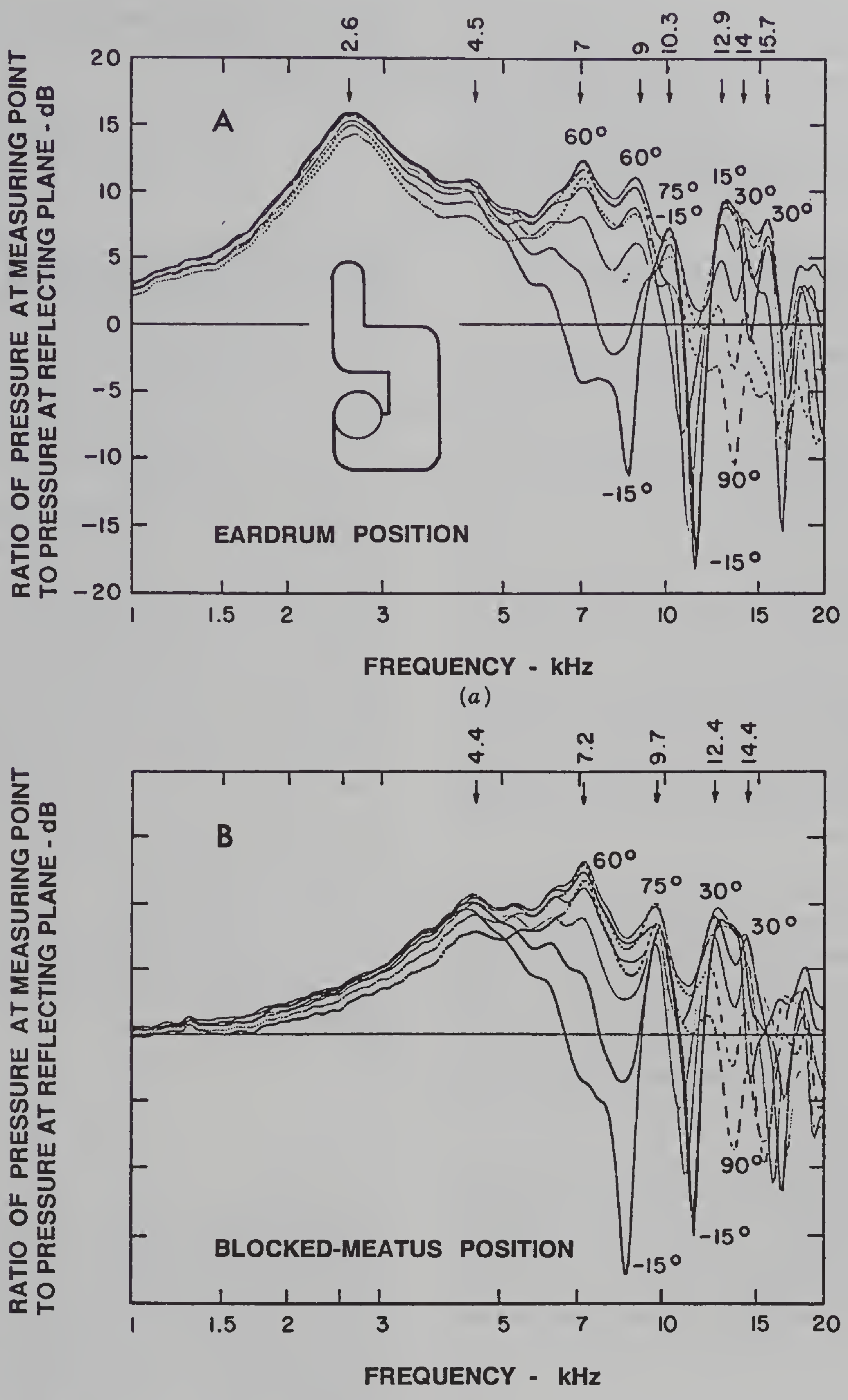
As can be seen, the five resonance frequencies in the model, in the absence of the canal [Panel (*b*)], are almost identical with the average values for human subjects shown in Fig. 3. Successful representations of the human concha and fossa based on other geometrical forms, such as the semicircle, can also be built.[3]

In the complete physical model featured in panel (*a*), the eardrum is represented by a condenser microphone and a two-branch acoustical network (NRC Network D) with characteristics that are compatible with the data given in Table 3. The canal is represented by a cylindrical cavity 7.5 mm in diameter with an adjusted length (≃25 mm) chosen to place the principal resonance frequency at 2.6 kHz in agreement with the average value for the human ear.[3] Evidently, the enfolding geometry at the junction between the "canal" and the "cavum" produces an effective canal length ($\lambda/4 \simeq 32$ mm) that is considerably greater than the length of the canal proper. The addition of the canal increases the number of resonances below 16 kHz from five to eight. However, at most frequencies within this range, the presence of the canal has little effect on the sound pressure distribution in the concha except in the vicinity of the ear canal entrance.

4.4 Response of Outer Ear Measured in Isolation

The families of response curves presented in Fig. 4 display the high-frequency characteristics of a particular physical model designated "IRE" (inclined rectangular cavity: Stage E) in terms of the response measured at the eardrum position [panel (*a*)] and at the center of a rigid airtight plug closing the canal entrance [panel (*b*)]. Such measurements are made with a special sound source designed to irradiate the human ear with clean progressive waves at grazing incidence and in virtual isolation from the head and torso.[3] All angles of incidence lie in the circumaural plane with 0° and 90° representing waves approaching the ear from the front and from above, respectively. As can be seen, the directionality measured at the "blocked meatus" is almost identical with the directionality at the eardrum.

The curves shown in panel (*b*), though taken with a physical model, could readily be included in a collection of response curves obtained with real human ears under comparable conditions.[3,17] In particular, in panel (*b*) as in the human data, with sound waves approaching from high angles of incidence (60°–90°), the response remains



(*a*)

(*b*)

Fig. 4 Response of physical model of human outer ear (IRE) with progressive wave source at eight angles of incidence in circumaural plane: (*a*) complete ear; (*b*) with ear canal closed at entrance (blocked meatus). Arrows indicate frequencies of normal modes. Shape of concha–fossa representation is shown in panel (*a*). (After Ref. 3.)

strong between 6 and 9 kHz but is weak between 11 and 15 kHz, whereas, with frontal incidence (−15° to +15°), the response is weak between 6 and 9 kHz but strong between 11 and 16 kHz. While these and other major acoustical features appear to be characteristic of human ears in general, the pattern of response for each individual ear is rich in detail and distinctive in character.[17]

4.5 Sound Pressure Measurements in the Ear Canal

The presence of longitudinal standing waves in the ear canal makes it necessary to measure sound pressure within 2 mm of the umbo if the input to the human eardrum is to be determined with an absolute accuracy of 1 dB at 10 kHz. It is also necessary to use a fine probe tube. Useful measurements can, however, be made at appreciable distances from the eardrum, as indicated in Fig. 1, provided that the transfer function between the point of measurement and the eardrum position can be measured or estimated.[2,6] Where a relative measure of sound pressure is sufficient, the quarter-wavelength minimum in the standing-wave pattern can be used as a position marker to ensure accurate placement and replacement of the probe microphone orifice in the canal. As noted earlier, measurements at the convenient and highly reproducible blocked-meatus position also have substantial value as indicators of directionality (see also Ref. 20).

At frequencies greater than 5 kHz, the wave system in the canal may be contaminated with evanescent transverse modes generated in the concha.[2,18] These modes are unlikely to be present at significant amplitudes, except in the vicinity of the entrance to the ear canal and at minima in the longitudinal standing-wave pattern, since they have high attenuation constants [e.g., 3 dB/mm for the (1, 1) mode at 20 kHz assuming a canal diameter of 7.5 mm].

The human ear canal can be treated as an acoustical horn in which the cross-sectional area is measured perpendicular to a curvilinear central axis that follows the shape of the canal. The length of this axis is appreciably greater than the superficial length of the canal.[1] This technique has also been used to interpret sound pressure measurements in the cat ear canal at frequencies as great as 30 kHz.[21]

5 DIFFUSE-FIELD RESPONSE AND OVERALL PERFORMANCE

By invoking the acoustical reciprocity principle, it can be shown that the diffuse-field response of the outer ear, p_d^2/p_f^2, is essentially determined by two impedances, Z_d and Z_a, as follows:

$$\frac{p_d^2}{p_f^2} = \left(\frac{c}{\pi\rho\nu^2}\right)\frac{\eta R_a|Z_d|^2}{|Z_a + Z_d|^2}. \tag{3}$$

In this expression, p_f^2 is the mean-square pressure in the diffuse sound field, p_d^2 is the mean-square pressure generated at the eardrum, Z_d is the load impedance presented to the outer ear by the eardrum, Z_a is the impedance presented to the eardrum by the outer ear (in essence, the radiation impedance), R_a is the real part of Z_a, and η is the acoustical radiation efficiency of the outer ear, while ν, ρ, and c are the frequency, density of air, and velocity of sound, respectively.[22]

Curve *A* in Fig. 5 is an estimated median diffuse-field response curve for the human ear based on synthesized response data for the mode-matched model ear featured in Fig. 4 and for replicas of several human ears, as well as measured values of Z_a and Z_d for the model ear.[3] This estimate is in broad agreement with other data.[23] The frontal incidence response (curve *B*) is well below the diffuse-field response between 5 and 10 kHz reflecting the relative insensitivity of the human ear in the horizontal plane at these frequencies (see Fig. 4). The intersubject differences in a diffuse sound field are generally much smaller than those found at any particular angle of incidence in the free field.

The overall performance of the outer ear as a "collector" of sound energy can also be quantified in terms of Z_d, Z_a, and η.[22,24,25] Hence, it can be shown that the maximum amount of sound power that can be received from a diffuse sound field is

$$W_d = \frac{\lambda^2}{4\pi}\frac{p_f^2}{\rho c}, \tag{4}$$

where λ is the wavelength, which is also the power that would flow into a perfectly absorbing sphere of cross-sectional area $\lambda^2/4\pi$ (radius $\lambda/2\pi$). This limit is attained where Z_a is the conjugate of Z_d and where there is no loss of energy during transmission to the eardrum (η = 1). More generally, the performance of any outer ear can be expressed in terms of its diffuse field absorption cross-section $A \leq \lambda^2/4\pi$.

Figure 6 shows the value of *A* as a function of frequency for the mode-matched model ear, neglecting radiation losses that are probably of the order of 1 dB.[3,24] Also shown is the value of *A* when the outer ear is eliminated and the "eardrum" (i.e., Z_d) is placed at the surface of a sphere representing the head. As can be seen,

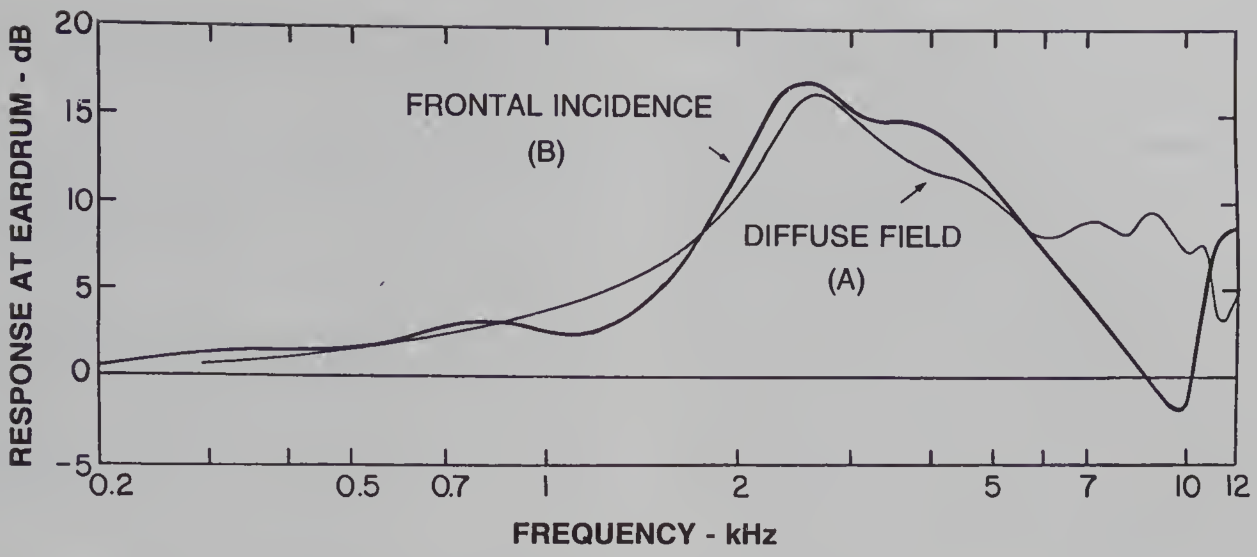


Fig. 5 Curve *A*: Estimated median transformation of sound pressure level from diffuse field to human eardrum. Curve *B*: Average transformation with frontal incidence: 0° azimuth in Fig. 2. (After Ref. 3.)

the absorption cross section of the human ear is not far below the theoretical limit at the principal resonance frequency (2.6 kHz) and at higher frequencies. Comprehensive measurements of Z_a made on outer ears excised from cats indicate that, in this species also, the value of A approaches the theoretical limit at the principal resonance frequency.[25]

6 EARS COUPLED TO EARPHONES

When sounds are presented to the outer ear through earphones, the free-field characteristics of the ear are replaced by the characteristics of an artificial system that are dependent on complex acoustical interactions between the individual earphone and the individual ear.

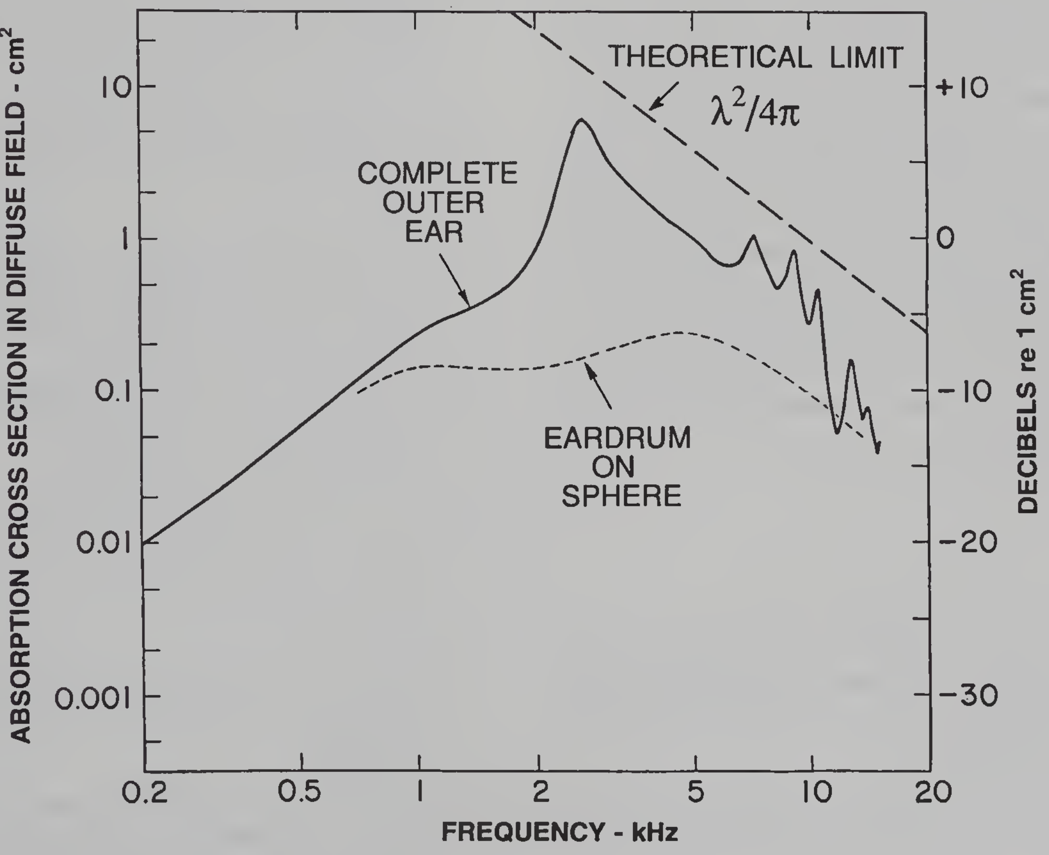


Fig. 6 Diffuse-field absorption cross section as a function of frequency for mode-matched physical model of human outer ear. (After Ref. 24.)

These interactions may include wave effects that are under poor control, especially at high frequencies.

Artificial ears, earphone couplers, and ear simulators reproduce selected acoustical characteristics of the human ear, with particular reference to the calibration of sound sources such as audiometric earphones and hearing aid receivers under standard conditions.[2,3,26,27] The simplest couplers are cylindrical cavities engaging with an earphone at one end and closed by a microphone at the other. In the NBS9A coupler, a single cavity represents the effective volume presented to a supra-aural earphone by the average human ear and the microphone provides an approximate measure of the sound pressure generated in the concha. In the more advanced IEC Artificial Ear, the main cavity, which includes the microphone, represents the volume of the concha when compressed by a supra-aural earphone, while auxiliary cavities represent the effective volume of the ear canal and eardrum, and the compliance of the pinna flange.[26] In occluded ear simulators, a cylindrical cavity, representing the inner portion of the ear canal, is terminated by a microphone surrounded by highly developed acoustical networks representing the impedance of the eardrum.[27] These simulators are well suited to the calibration of insert earphones and can be embodied in realistic "manikins" that include replicas or representations of the human concha, pinna flange, head, and torso.[9]

The limitations inherent in coupler methods of calibration are avoided when earphones are calibrated on real ears. The sound pressures generated by plane waves approaching the head from one or many directions in a free-field are first measured with a probe microphone, at a selected position in the ear canal, to create a set of *head-related transfer functions* (HRTFs).[14,20] These functions are then compared with the sound pressure generated at the same canal position by sounds presented through the earphone. Ideally, separate sets of HRTF and earphone response functions are created for each listener. Recently, it has been shown that some earphones can be successfully calibrated with measurements made at the blocked-meatus position shown in Fig. 1.[28]

REFERENCES

1. M. R. Stinson and B. W. Lawton, "Specification of the Geometry of the Human Ear Canal for the Prediction of Sound-Pressure Level Distribution," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 2492–2503.
2. E. A. G. Shaw, "The External Ear," in W. D. Keidel and W. D. Neff (Eds.), *Handbook of Sensory Physiology*, Vol. V/1, Springer-Verlag, Berlin, 1974, pp. 455–490.
3. E. A. G. Shaw, "The Acoustics of the External Ear," in G. A. Studebaker and I. Hochberg (Eds.), *Acoustical Factors Affecting Hearing Aid Performance*, University Park Press, Baltimore, 1980, pp. 109–124.
4. R. Teranishi and E. A. G. Shaw, "External-Ear Acoustic Models with Simple Geometry," *J. Acoust. Soc. Am.*, Vol. 44, 1968, pp. 257–263.
5. B. Kruger and R. J. Ruben, "The Acoustic Properties of the Infant Ear," *Acta Otolaryngol. (Stockh.)* Vol. 103, 1987, pp. 578–585.
6. E. A. G. Shaw, "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1848–1861.
7. E. A. G. Shaw and M. M. Vaillancourt, "Transformation of Sound-Pressure Level from the Free Field to the Eardrum Presented in Numerical Form," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 1120–1123.
8. S. Mehrgardt and V. Mellert, "Transformation Characteristics of the External Human Ear," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 1567–1576.
9. M. D. Burkhard and R. M. Sachs, "Anthropometric Manikin for Acoustic Research," *J. Acoust. Soc. Am.*, Vol. 58, 1975, pp. 214–222.
10. J. C. Middlebrooks, J. C. Makous and D. M. Green, "Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 89–108.
11. M. B. Calford and J. D. Pettigrew, "Frequency Dependence of Directional Amplification at the Cat's Pinna," *Hearing Res.*, Vol. 14, 1984, pp. 13–19.
12. A. D. Musicant, J. C. K. Chan, and J. E. Hind, "Direction-dependent Spectral Properties of Cat External Ear: New Data and Cross-Species Comparisons," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 757–781.
13. G. F. Kuhn, "Model for the Interaural Time Differences in the Azimuthal Plane," *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 157–167.
14. J. Blauert, *Spatial Hearing*, MIT Press, Cambridge, 1983.
15. J. C. Middlebrooks and D. M. Green, "Directional Dependence of Interaural Envelope Delays," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 2149–2162.
16. M. R. Stinson, "Revision of Estimates of Acoustic Energy Reflectance at the Human Eardrum," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 1773–1778.
17. E. A. G. Shaw, "External Ear Response and Sound Localization," in R. W. Gatehouse (Ed.), *Localization of Sound: Theory and Applications*, Amphora Press, Groton, CT, 1982, pp. 30–31.
18. R. D. Rabbitt and M. T. Friedrich, "Ear Canal Cross-Sectional Pressure Distributions: Mathematical Analysis and Computation," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 2379–2390.
19. M. R. Stinson, E. A. G. Shaw, and B. W. Lawton, "Estimation of Acoustical Energy Reflectance at the Eardrum from Measurements of Pressure Distribution in the Human Ear Canal," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 766–773.
20. H. Møller, M. F. Sorenson, D. Hammershoi, and C. B.

Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.*, Vol. 43, 1995, pp. 300–321.

21. M. R. Stinson and S. M. Khanna, "Sound Propagation in the Ear Canal and Coupling to the Eardrum, with Measurements on Model Systems," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 2481–2491.
22. E. A. G. Shaw, "Diffuse Field Response, Receiver Impedance, and the Acoustical Reciprocity Principle," *J. Acoust. Soc. Am.*, Vol. 84, 1988, pp. 2284–2287.
23. G. F. Kuhn, "The Pressure Transformation from a Diffuse Sound Field to the External Ear and to the Body and Head Surface," *J. Acoust. Soc. Am.*, Vol. 65, 1979, pp. 991–1000.
24. E. A. G. Shaw, "1979 Rayleigh Medal Lecture: The Elusive Connection," in R. W. Gatehouse (Ed.), *Localization of Sound: Theory and Applications*, Amphora Press, Groton, CT, 1982, pp. 13–29.
25. J. J. Roskowski, L. H. Carney, and W. T. Peake, "The Radiation Impedance of the External Ear of Cat: Measurements and Applications," *J. Acoust. Soc. Am.*, Vol. 84, 1988, pp. 1695–1708.
26. *American National Standard Method for Coupler Calibration of Earphones*, S3.7-1995, Acoustical Society of America, New York, 1995.
27. *American National Standard for an Occluded Ear Simulator*, S3-25-1989, Acoustical Society of America, New York, 1995.
28. H. Møller, D. Hammershoi, C. B. Jensen, and M. F. Sørensen, "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.*, Vol. 43, 1995, pp. 203–216.

106

ACOUSTIC PROPERTIES OF THE MIDDLE EAR

WILLIAM T. PEAKE AND JOHN J. ROSOWSKI

1 INTRODUCTION

The middle ear couples acoustic signals from the external ear canal to the inner ear. Measurements of its acoustic properties include acoustic input impedance and transfer ratios. As a signal transmission system the middle ear is essentially linear; the pressure transfer ratio is bandpass with a gain of 20–30 dB in the midband. The gain can be reduced by contraction of either of two small muscles or by static-pressure difference between the middle-ear air space and the ear canal. Acoustic measures are useful for diagnosis of some hearing disorders.

2 DESCRIPTION OF THE MIDDLE EAR

2.1 Structure

The middle ear consists of the tympanic membrane, the ossicular chain with its attached ligaments and muscles, and the air-filled cavity that contains the ossicles (see Fig. 1*a*). The tympanic membrane is the boundary between the external ear and the middle ear; the oval and the round windows of the cochlea are boundaries between the middle ear and the inner ear. The ossicles link motion of the tympanic membrane to the inner ear through motion of the stapes in the oval window. This description is appropriate for all terrestrial vertebrate species except those with no tympanic membrane (e.g., salamanders[1] and snakes[2]).

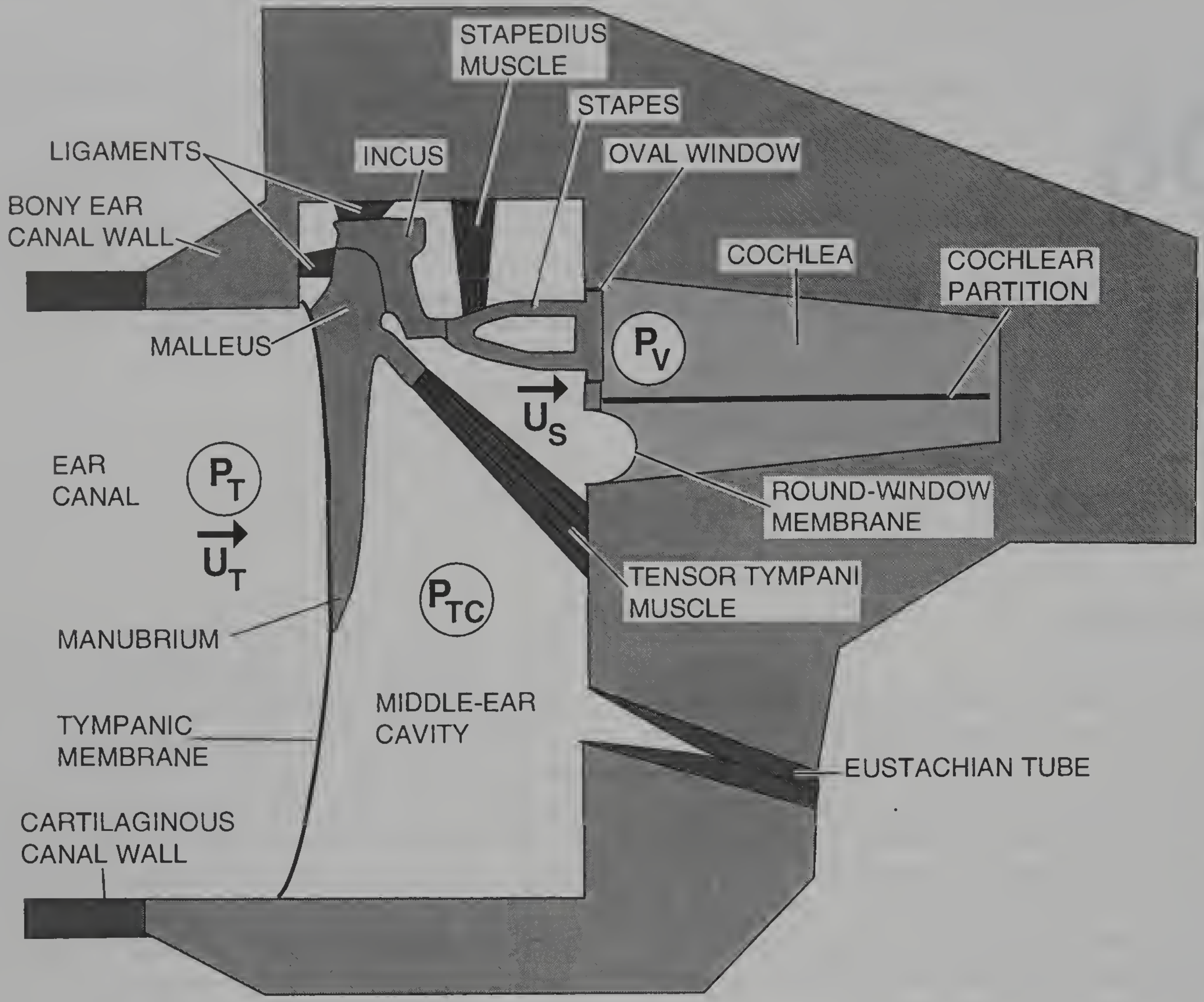
In mammals the ossicular chain consists of a linkage of three bones, the malleus (hammer), the incus (anvil), and the stapes (stirrup). Although the configuration of these bones varies among species,[1,3,4] some features occur regularly. An elongated process of the malleus (the manubrium, or handle of the hammer) is embedded between layers of the tympanic membrane. The malleus and the incus are connected to the bony walls of the middle ear by ligaments and to each other through a joint. The incus connects through another joint to the stapes, the inner face of which (the stapes footplate) is attached by an annular ligament in the oval window of the cochlea. Two muscles attach to the ossicles, the tensor-tympani muscle to the malleus and the stapedius to the stapes. The configuration of the middle-ear cavity varies among mammalian species[5,6]; in the human ear the main (tympanic) cavity is connected through a narrow region (the aditus) to a second cavity (the antrum), which is coupled to numerous smaller cavities that make up the "aerated" mastoid region of the temporal bone.[7] The eustachian tube, which connects the tympanic cavity to the nasopharynx, is usually closed, but it can be opened by contraction of attached muscles.[8,9]

2.2 Function

The function of the middle ear is to transmit acoustic signals from the ear canal to the cochlea. This signal coupling takes place primarily through the action of the tympanic membrane (TM) and the ossicular chain. For the human ear, except for the upper part of the audio-frequency range (i.e., above 5 kHz), the transverse dimensions of the ear canal are small relative to sound wavelengths and the sound pressure is approximately constant across the TM.[10,11] In this case the acoustic interactions at the middle-ear input (i.e., at the tympanic membrane) can be described in terms of two variables, the sound pressure P_T at the TM and the TM volume velocity U_T. The motion of the TM and the attached malleus is coupled through the three ossicles. The incudomallealar and

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

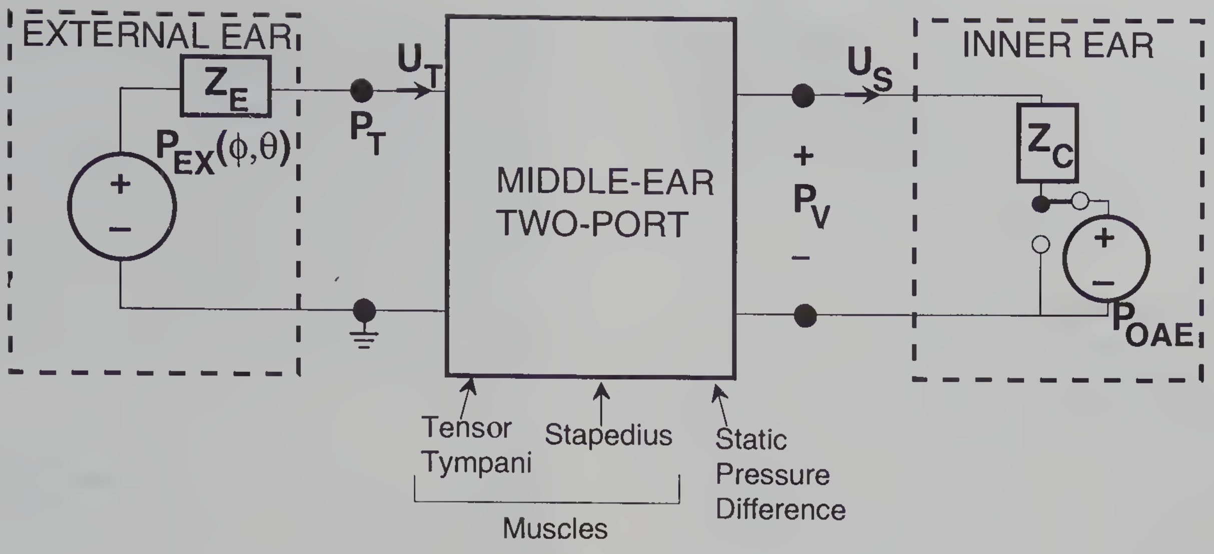

STAPEDIUS MUSCLE
STAPES
LIGAMENTS
INCUS
OVAL WINDOW
BONY EAR CANAL WALL
COCHLEA
COCHLEAR PARTITION
MALLEUS
P_V
U_S
EAR CANAL
P_T
U_T
ROUND-WINDOW MEMBRANE
TENSOR TYMPANI MUSCLE
P_{TC}
MANUBRIUM
MIDDLE-EAR CAVITY
TYMPANIC MEMBRANE
EUSTACHIAN TUBE
CARTILAGINOUS CANAL WALL

(*a*)

Fig. 1 (*a*) Schematic representation of the structures of the mammalian middle ear with acoustic variables indicated that will be used in the text. The five labeled acoustic variables in the figure represent complex amplitudes of sinusoidal functions of time t with frequency f. For instance, the pressure in the ear canal $p_T(t) = |P_T| \cos(2\pi f t + \angle P_T)$. Other variables are U_T = the volume velocity at the tympanic membrane, P_{TC} = the sound pressure in the tympanic cavity, U_S = the volume velocity of the stapes footplate, and P_V = the sound pressure in the inner-ear vestibule just inside the stapes footplate. Components of the middle-ear cavity configuration are not represented in this picture. (*b*) An analog-circuit representation of the signal-transmission function of the middle ear. The "impedance analogy" is used: that is, volume velocity is analogous to current and sound pressure is analogous to voltage. The middle ear is represented by a two-port network with input variables P_T, U_T and output variables P_V, U_S. External and inner ears are represented by Thévenin equivalents. The equivalent external source $P_{\text{EX}}(\phi, \theta)$ is a function of the sound propagation direction (ϕ, θ) relative to the head; Z_E is the output impedance of the external ear; Z_C is the input impedance of the cochlea. A Thévenin equivalent source for otoacoustic emissions is indicated as P_{OAE}; if the switch with this source is thrown from the position drawn here, this source is removed, which produces a good approximation of most middle-ear behavior. Influences of other variables on the middle ear are indicated by arrows below the two-port.

incudostapedial joints are thought to introduce little relative motion in response to sound,[12–16] but they flex appreciably during muscle contraction[17] and with large static pressures across the tympanic membrane.[18] The approximately pistonlike motion of the stapes into the fluid of the cochlea, which can be described by a volume velocity U_S, generates a sound pressure in the vestibule P_V.

In the classical view the lever mechanism produced by the TM with rigid-body rotational motion of the malleus and incus increases the sound pressure from external ear



(*b*)

Fig. 1 (*Continued*)

to inner ear.[19,20,21] Recent measurements of displacements along the manubrium of the malleus in response to sound indicate that the motion is more complicated than pure rotation, at least for some conditions, and the manubrium does not move as a rigid body at higher frequencies.[22,23] Thus, it seems clear that mechanisms other than a rigid lever are involved.[24]

The acoustic transmission of the middle ear can be represented by an acoustic two-port (Fig. 1*b*), in which the input terminal pair is connected to the output of the external ear and the output terminal pair is connected to the cochlea (Fig. 1*b*). Measurements (and theory[25]) indicate that the acoustic input impedance of the inner ear Z_C is approximately resistive at least in the mid-frequency range.[16,26,27] In normal operation transmission of acoustic signals through the middle ear is altered (usually reduced) by contractions of the middle-ear muscles[28] and by static pressure difference across the tympanic membrane,[29] which occurs either because of changes in ambient pressure (e.g., as a result of a change in altitude) or because of exchange of gas between the middle-ear cavity and its soft-tissue lining.[30] These static pressure differences are normally removed by momentary openings of the eustachian tube, which equalize the middle ear and ambient pressures. Pathological conditions in the middle ear also produce reductions in transmission through the ossicular chain. For instance, with middle-ear inflammation (otitis media) fluid builds up in the middle-ear cavity that can alter the mobility of the TM and ossicles[31,32]; in otosclerosis abnormal growth of bone around the footplate of the stapes can stiffen the suspension of the stapes so that less signal is coupled to the cochlea.[33] Reduced hearing sensitivity that results from interference with the mechanical operation of the middle ear is called *conductive hearing loss*. Such hearing deficits may be alleviated either through external amplification (hearing aids) or through medical and surgical procedures.[34–36]

In recent years measurements have shown that a variety of acoustic signals can be recorded in the external ear that seem to originate in the cochlea (P_{OAE} in Fig. 1*b*). These signals, which originate in the inner ear, referred to collectively as *otoacoustic emissions* (OAE),[37–40] are coupled to the external ear by the middle ear. As OAEs are relatively small and apparently have little effect on inward transmission through the middle ear, they will not be considered further. That is, we will assume that the switch in Fig. 1*b* is thrown so as too disconnect P_{OAE} from the circuit and the ear canal.

3 TERMINAL PERFORMANCE OF THE MIDDLE EAR

3.1 Linearity

In response to sound stimuli (and in the absence of reflex contractions of the middle-ear muscles) the middle ear operates approximately as a linear acoustic system up to very high sound pressure levels (SPL). In ears of anesthetized cats response amplitudes are proportional to stimulus amplitude for tonal stimuli with P_T up to 140 dB SPL.[13,26,41,42]

One kind of nonlinear behavior of the middle ear is commonly experienced when static pressure builds up in the middle ear. The reduction in acoustic sensitivity that accompanies static pressure differences of the order of 10^{-2} atmospheres (e.g., during aircraft ascent or descent) demonstrates that these large pressures (relative to normal *sound* pressures) can produce substantial variations from linear behavior. Another kind of nonlinearity has been demonstrated in which the input impedance varies with the level of the input sound pressure.[43] These rather small deviations from linearity, which occur in narrow frequency bands at low stimulus levels, are apparently generated by cochlear processes and are called *stimulus-frequency otoacoustic emissions*.

3.2 Acoustic Input Impedance

Measurements of acoustic impedance (or admittance) at the TM have been used to assess the acoustic behavior of the middle ear for scientific,[44] medical,[33] and engineering[45] purposes. Measurements in live humans are complicated by uncertainty about the representation of the coupling space between the measurement site and the tympanic membrane, which has a large effect at the higher frequencies.[46] In human cadaver ears measurements can be made near to the TM, the coupling space can be assessed more accurately, and results have been extended to somewhat higher frequencies (5 kHz[47]). In ears of anesthetized animals measurements have been reported to above 10 kHz.[44,48,49] Measurements in human ears also extend to higher frequencies, but the location of the measurements relative to the tympanic membrane is not known accurately.[50–52]

Impedance measurements in human ears (Fig. 2*a*) have some consistent features. At low frequencies (i.e., below 0.3 kHz) the angle is near −0.25 periods and the magnitude is inversely proportional to frequency; this is the behavior of an acoustic compliance (or mechanical spring). At higher frequencies (1–3 kHz) the impedance angle is near zero and the magnitude does not show large changes with frequency, which is resistance-like behavior. Interear variations in magnitude of somewhat less than a factor of 10 are consistent with measurements over larger population of ears at low frequencies.[53–55] These features are also generally seen in measurements from other mammalian species (Fig. 2*b*).

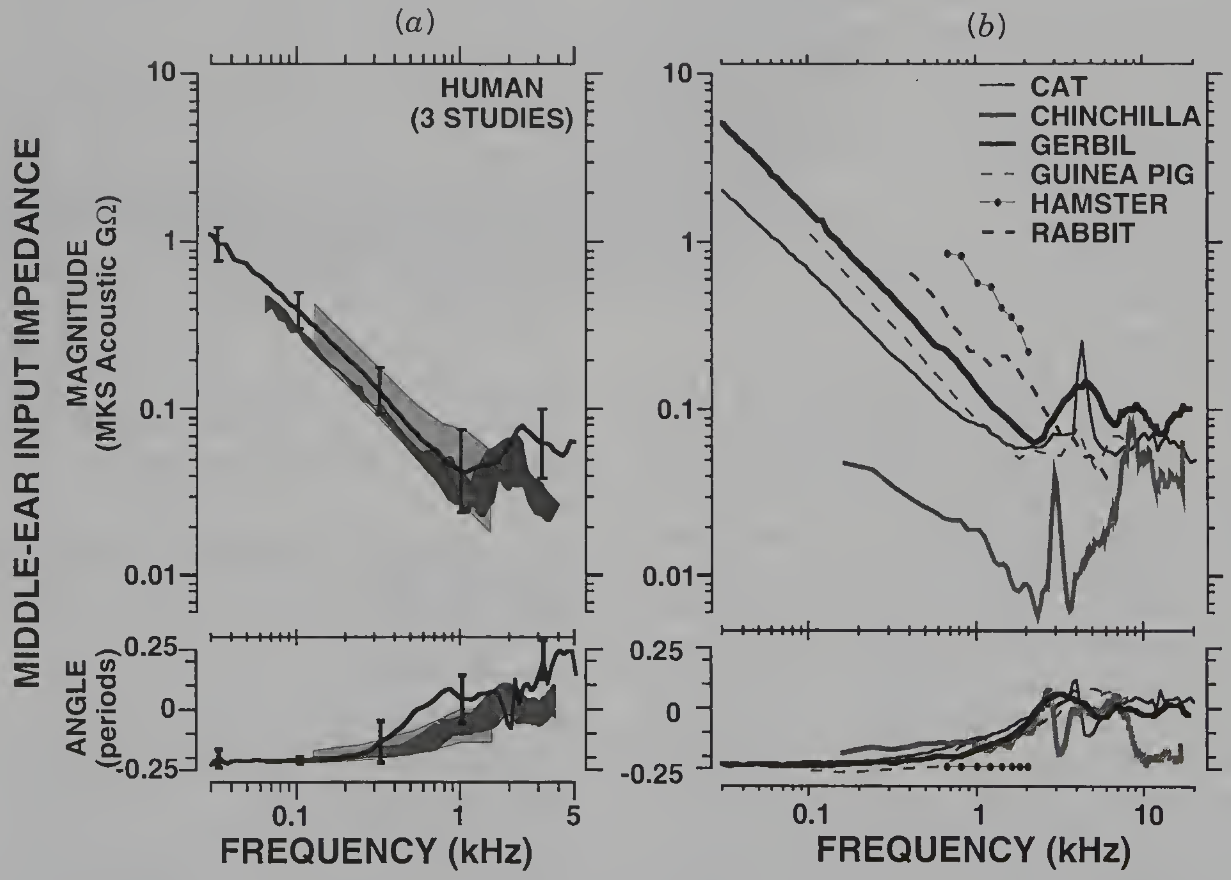


Fig. 2 Measurements of the acoustic input-impedance $Z_T = P_T/U_T$ at the tympanic membrane of (*a*) human ears and of (*b*) some other mammals. Magnitudes and angles of the measured impedances are plotted vs. frequency. The vertical scale for the magnitude plots is given in MKS acoustic $G\Omega = 10^9 \text{Pa-s/m}^3$. For the angle plots the indicated scale is in periods where 1 period = 2π radians = 360°. (*a*) A summary of human ear measurements from three studies. The lighter shaded area represents the 10th through the 90th percentile ranges of 33 subjects with normal hearing.[33] The darker shaded area is the range for 4 young males with normal hearing.[46] The solid line (with error bars) represents the median and standard deviation from measurements in 15 cadaver ears.[47] (*b*) Measurements of acoustic input impedance of the middle ear in six mammalian species. For each species average magnitudes and phase angles are plotted.[44]

The acoustic impedance of ears with pathological conditions that cause conductive hearing losses can deviate substantially from normal.[33,35] However, in spite of numerous investigations of the clinical application of such results,[56–58] measurements of impedance magnitude are not now widely used in diagnosis of middle ear pathologies, partly because their ability to define the nature and location of the abnormality is too limited and partly because surgeons find other diagnostic methods preferable.

3.3 Transfer Functions

Two kinds of middle-ear transfer ratio are shown in Fig. 3. In Fig. 3*a* the ratio of output volume velocity U_S to input pressure P_T is shown. At low frequencies this transfer admittance's magnitude increases proportionally with frequency and its angle leads the applied pressure by about 0.25 periods. This result indicates that the middle-ear structures are displaced in phase with each other and in phase with the input pressure, and suggests that motion is primarily determined by elastic forces. At the highest frequencies the angle of the transfer function decreases, generally becoming more negative than -0.25 periods; the total change in angle with frequency of more than 0.5 periods indicates that an analytical representation of the middle-ear transmission function must be of higher order than second; that is, if a wide frequency range is considered, the middle ear is more complex than a simple harmonic oscillator.

The input–output pressure ratio (Fig. 3*b*) shows that in the midfrequency range the middle ear produces a pressure gain of 20–30 dB. Presumably the mechanism responsible is primarily the ratio of TM area to stapes-

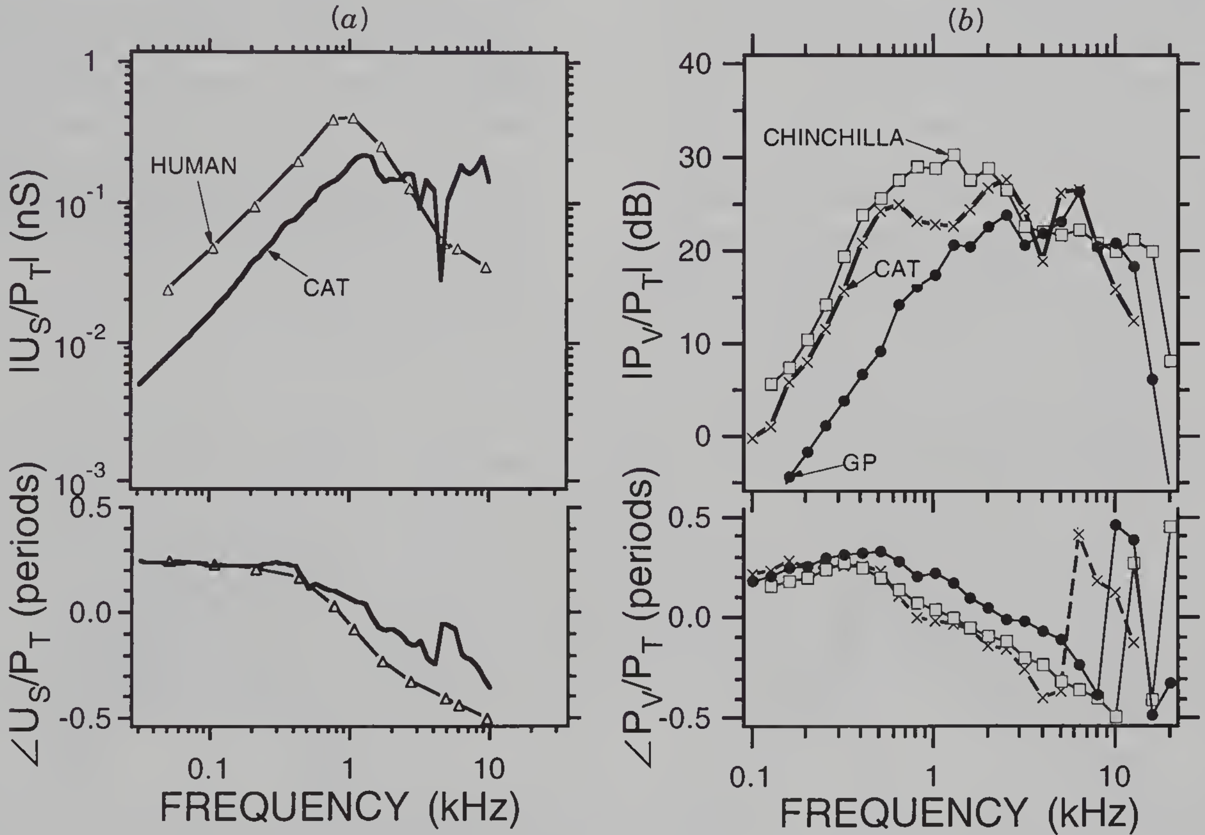


Fig. 3 Measurements of transfer functions of middle ears. In each case both magnitudes (upper) and angles (lower) are plotted versus frequency. (*a*) The acoustic transfer admittance U_S/P_T from human and cat ears. The vertical scale is acoustic nanoSiemens (i.e., 1 nS = 10^{-9}m^3/Pa-s). The cat measurements were made on 25 anesthetized animals with observation of the stapes motion through a microscope[13]; corrections were made for the effects of opening the middle-ear cavities.[59] The human results[81] were obtained on 20 cadaver ears through acoustic measurement of volume velocity of the round window, which is assumed to equal the volume velocity of the stapes.[82] (*b*) Averaged measurements of the pressure transfer ratio P_V/P_T for cat (N = 6), guinea pig (GP) (N = 5), and chinchilla (N = 4). In each case the sound pressure in the vestibule of the inner ear was determined by a pressure transducer that was introduced through a hole drilled into the bone of the inner ear.[15] Middle-ear cavities are "intact." In the angle plot (lower, right) each of the plotted values jumps at a high frequency by about 1 period because the plotted angles are restricted to the ±0.5 period range. (From Ref. 15. Used with permission.)

footplate area, which is in the range 26–30 dB for many species (see Ref. 4, Fig. 6.15). The frequency dependence of the pressure gain demonstrates that the TM-ossicular chain does not act as an ideal transformer.[4,24]

From measurements of input impedance and transfer ratios it is possible to determine a power transmission efficiency for the middle ear. Available measurements on cat ears indicate that over the audio-frequency range the efficiency is a maximum of about 50%; that is, at least half the power absorbed into the middle ear is not delivered to the inner ear.[59]

3.4 Alterations in Middle-Ear Transmission

Static Air Pressure Differences between static pressures in the ear canal and in the middle-ear cavity can change the acoustic input impedance, forward sound transmission, and outward transmission of cochlear acoustic emissions. Pressure differences of 10 cm of H_2O (0.01 atm = 1 kPa) produce detectable changes in humans ears[60–63] and in other mammals (e.g., Fig. 4*a* and *b*). Positive pressures produce reductions that increase with the size of the pressure difference and are generally larger for the lower frequencies. Negative pressures have a more complex pattern with some pressures producing an increase in admittance for higher frequencies. The lack of equality of the input admittance change (Fig. 4*a*) and the transmission change (Fig. 4*b*) indicates that the stiffening caused by static pressures changes the coupling of TM velocity to stapes velocity.

The effect of static pressure is the basis of a clinical diagnostic procedure. Measurements of the input admittance magnitude (usually at a low frequency, such as 226 Hz) as a function of the static pressure (e.g., ±0.2 kPa) applied in the ear canal are called tympanograms.[57,58] In normal ears tympanograms have a sharp maximum (in admittance magnitude) when the ear canal static pressure equals that in the middle ear and the magnitude decreases for positive or negative pressures differences. Deviations from this pattern are evidence for conductive abnormalities in the middle ear. For instance, flat tympanograms (i.e., an input admittance magnitude that is relatively insensitive to static-pressure changes) indicate the presence of fluid in the middle ear.[53]

Middle-Ear Muscle Contractions The middle ear muscles contract under a variety of circumstances including high intensity sound stimulation (the "acoustic reflex"), before and during vocalization, after tactile stimulation to the skin of the head, and during general body motion.[64] The stapedius muscle has been more completely characterized than the tensor tympani, but the effects of their contractions appear to be similar, at least in rabbit, cat, and guinea pig.[29,65,66]

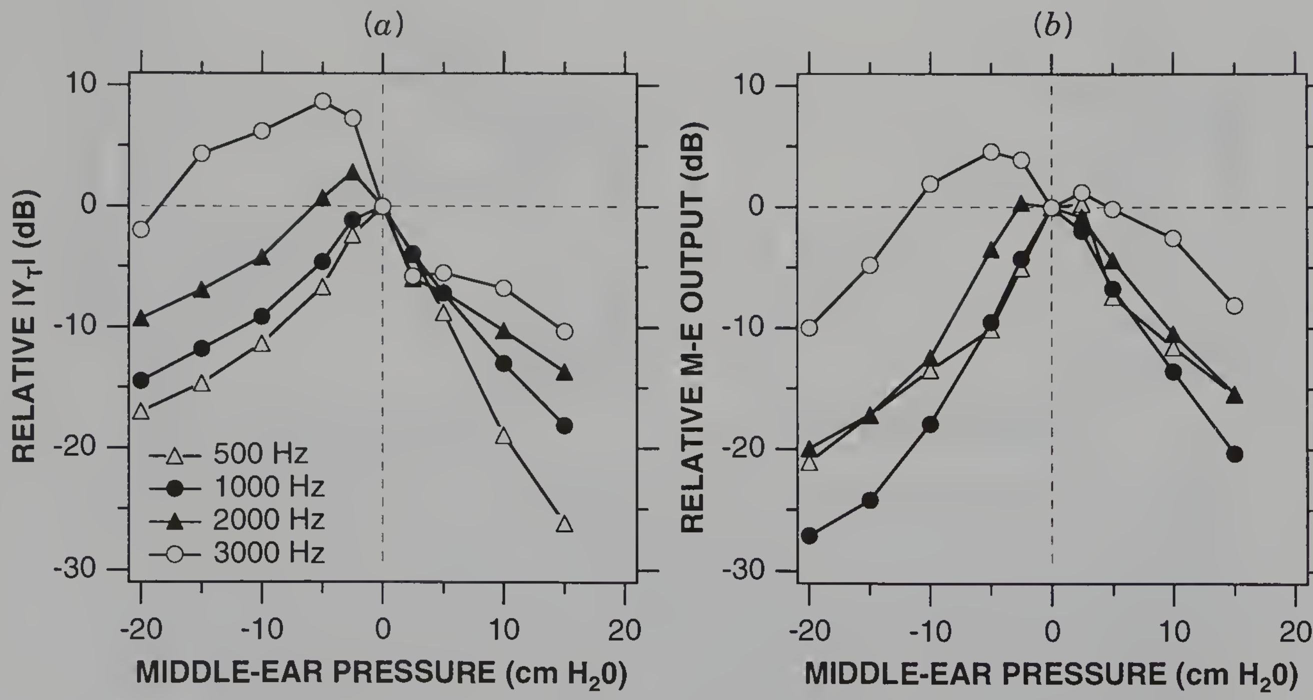


Fig. 4 Effects of static pressure difference across the tympanic membrane on the magnitudes of both the acoustic input admittance $Y_T = U_T/P_T = (Z_T)^{-1}$ and the transmission of a middle ear of cat. The abscissa in both plots is the static pressure in the middle-ear airspace P_{ME} relative to ambient. Measurements are shown for four frequencies. In (*a*) the ordinate is $20\log_{10}|Y_T(P_{\mathrm{ME}})|/|Y_T(0)|$. The transmission changes in (*b*) were measured in the cochlear electric potential recorded in response to tonal stimuli. (Data from Ref. 12; similar, but more extensive transmission change results from cat ears are reported in Ref. 19.)

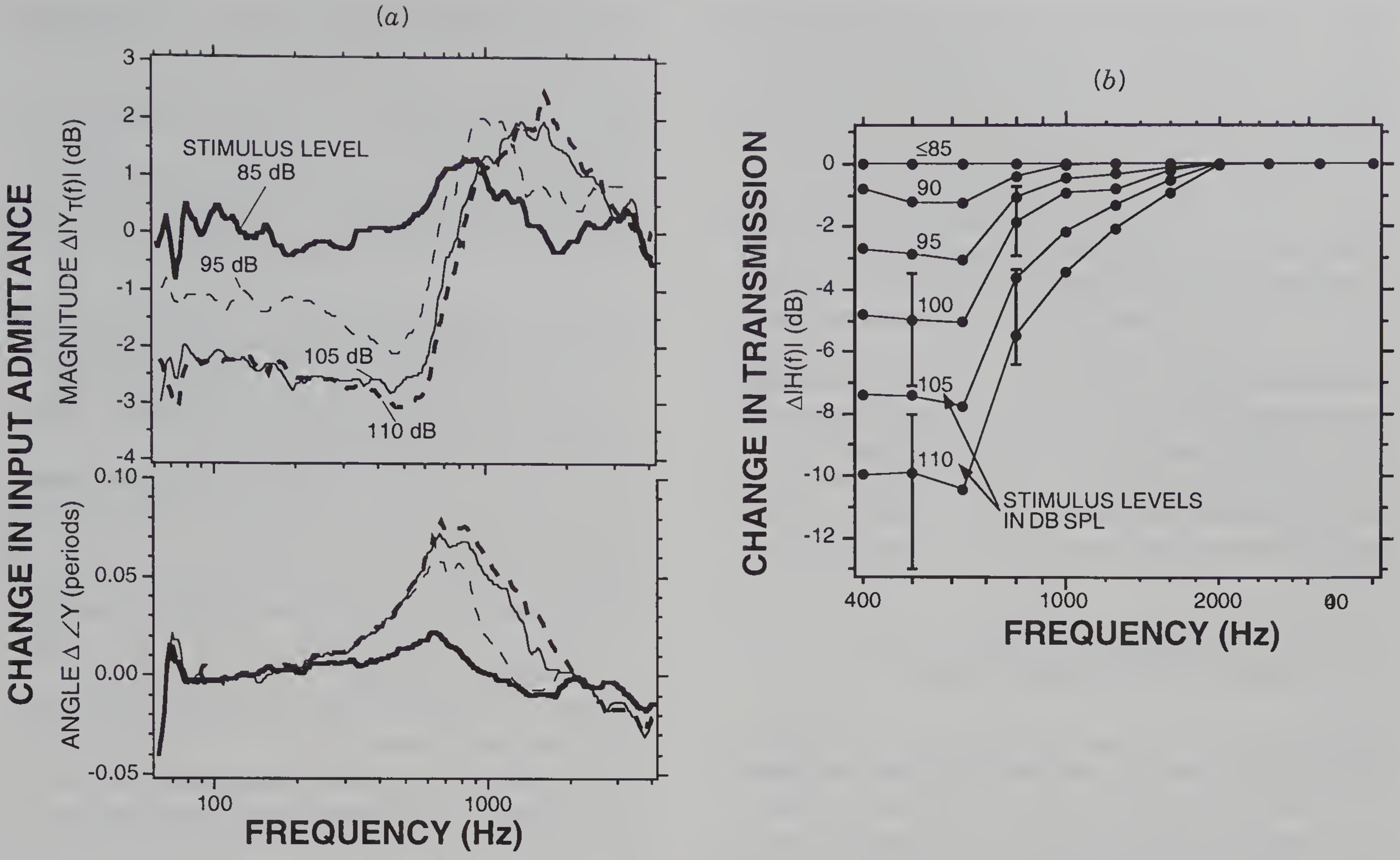


Fig. 5 Effects of the contralateral acoustic reflex on both the acoustic input admittance and the transmission of the human middle ear. The acoustic reflex was produced by relatively intense bursts of 2–4 kHz bandpass noise delivered (through an earphone) to the contralateral ear. The duration of each noise burst was 0.65 s with the impedance measurements made during the last 0.25 s, when the relex contraction appeared to be approximately constant. (*a*) Reflex-induced changes in the magnitude and angle of the acoustic input admittance vs. frequency with the level of the reflex-producing noise as a parameter. The $\Delta|Y_T|$ ordinate is plotted as $20\log_{10}|Y_T(during\ reflex)|/|Y_T(without\ reflex)|$. From one subject (of 4) whose reflex changes were nearest to the mean. (*b*) The average change in the magnitude of the middle-ear transmission is shown for the four subjects. These measurements (made at 11 frequencies) are based on a psychophysical method in which the rapid change with input level of the phase of a perceptual combination tone, $2F_1–F_2$, is used.[77,83] (From Ref. 77, pp. 196, 209. Used with permission.)

In human ears changes in the middle-ear acoustic behavior following brief high-intensity sound stimulation are primarily the result of reflex contraction of the stapedius.[28] These changes can be studied by applying the reflex-evoking sound to one ear while measuring changes in the opposite ear. Results of such measurements are shown in Fig. 5. In Fig. 5*a* the changes in magnitude and angle of the input admittance are plotted. For frequencies below 0.5 kHz the admittance magnitude decreases by a factor that is approximately independent of frequency and the angle is roughly unchanged. These results indicate that the compliance of the ear is decreased by the reflex. A change is seen for levels of the evoking stimulus greater than 85 dB SPL, and the effect ceases growing with level between 105 and 110 dB. At higher frequencies (1–3 kHz) the admittance magnitude is increased during the reflex.

The corresponding changes in transmission (Fig. 5*b*) are reductions in magnitude at low frequencies again with approximately constant (with frequency) changes for frequencies below 0.63 kHz. The transmission changes appear first for stimulus levels above 85 dB SPL as do the admittance changes, but the transmission continues to decrease with increasing stimulus level up to the highest level (110 dB).

The action of the stapedius contraction is to displace the head of the stapes in a direction normal to its motion in response to sound. Measurements in cat show that transmission decreases of 10 dB (as in Fig. 5*b*) result from a displacement of 40 μm. With muscle contraction the stapes head slides parallel to the incudo-stapedial joint surface and does not produce a comparable motion of the incus.[17]

Various consequences of the acoustic reflex have been shown: (1) The absence of the reflex, in patients with Bell's palsy, can have detrimental effects on the discrimination of speech sounds presented at high levels.[67] (2) The reflex can reduce the loss of hearing sensitivity that occurs from prolonged high-intensity sound stimulation.[68] (3) Activation of the stapedius muscle can have a large effect (e.g., 40 dB) in reducing masking by low-frequency sound of high-frequency stimuli.[69,70] (4) In clinical practice detection of reflex-evoked changes in acoustic impedance provides objective evidence of normal operation of the lower part of the central auditory system.[71]

4 MODELS OF THE MIDDLE EAR

Knowledge of the acousto-mechanical relationships involved in the middle ear's operation have been represented in network models, usually presented as electric-circuit analogs.[66,72–76] The effects of the muscles may be represented by changes of impedances when the muscle contracts.[66]

Circuit models of this kind can be an efficient summary of our knowledge of the middle ear. For instance, the model of Kringlebotn[76] can be used to represent the input impedance and middle-ear transfer function of normal and pathological human ears,[74] and transmission measurements. This kind of model can also explain the observation (Fig. 5) that the effect of the stapedius reflex on the input admittance saturates with increasing stimulus, even as the transmission continues to decrease.[77] However, we must remember that we have built assumptions into the models. For instance, most models assume that only one variable is required to represent the motion of the malleus. If, as has been reported,[22] the components of mallear motion or the location of the rotational axis changes (e.g., with frequency or with contraction of the muscle), or the manubrium bends,[23,78] the model must be modified to include these features. Another kind of behavior such models do *not* represent is the effects of coupling to the cochlea through mechanisms other than the motion of the ossicular chain.[79,80]

REFERENCES

1. O. W. Henson, "Comparative Anatomy of the Middle Ear," in W. D. Keidel and W. D. Neff (Eds.), *Handbook of Sensory Physiology, Vol. V/1, Auditory System. Anatomy. Physiology (Ear)*, Springer-Verlag, New York, 1974, pp. 39–110.
2. E. G. Wever, *The Reptile Ear*, Princeton University Press, Princeton, NJ, 1979.
3. G. Fleischer, "Evolutionary Principles of the Mammalian Middle Ear," in *Advances in Anatomy Embryology and Cell Biology*, Hill-Fay Associates, Winnetka, IL, 1978, pp. 3–69.
4. J. J. Rosowski, "Outer and Middle Ears," in R. R. Fay and A. N. Popper (Eds.), *Comparative Hearing: Mammals*, Springer-Verlag, New York, 1994, pp. 172–247.
5. I. Kirikae, *The Structure and Function of the Middle Ear*, University of Tokyo Press, Tokyo, 1960.
6. M. J. Novacek, "Aspects of the Problem of Variation, Origin and Evolution of the Eutherian Auditory Bulla," *Mammal Rev.*, Vol. 7, 1977, pp. 131–150.
7. B. J. Anson and J. A. Donaldson, *The Surgical Anatomy of the Temporal Bone and Ear*, Saunders, Philadelphia, 1967.
8. J. Sadé (Ed.), *The Eustachian Tube*, Kugler, Amsterdam/Berkeley, 1978.
9. I. Honjo, *Eustachian Tube and Middle Ear Diseases*, Springer-Verlag, Tokyo, 1988.
10. M. R. Stinson, "The Spatial Distribution of Sound Pressure within Scaled Replicas of the Human Ear Canal," *J. Acoust. Soc. Am.*, Vol. 78, 1981, pp. 596–1602.
11. M. R. Stinson and S. M. Khanna, "Sound Propagation in the Ear Canal and Coupling to the Eardrum, with Measurements on Model Systems," *J. Acoust. Soc. Am.*, Vol. 85, 1989, 2481–2491.
12. A. R. Møller, "Transfer Function of the Middle Ear," *J. Acoust. Soc. Am.*, Vol. 35, 1963, pp. 1526–1534.
13. J. J. Guinan, Jr. and W. T. Peake, "Middle-Ear Transmission in the Anesthetized Cat," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 1237–1261.
14. W. R. Rhode, "Some Observations on Cochlear Mechanics," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 158–176.
15. L. Décory, "Origine des Différences Interspecifiques de Susceptibilité an Bruit," Thése de Doctorat de l'Université Bordeaux II, France, 1989.
16. M. A. Ruggero, N. C. Rich, L. Robles, and B. G. Shivapuja, "Middle-Ear Response in the Chinchilla and Its Relationship to Mechanics at the Base of the Cochlea," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 1612–1629.
17. X. D. Pang and W. T. Peake, "How Do Contractions of the Stapedius Muscle Alter the Acoustic Properties of the Ear?" in J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Ed.), *Peripheral Auditory Mechanisms*, Springer-Verlag, New York, 1986.
18. K. B. Hüttenbrink, "The Mechanics of the Middle-Ear at Static Air Pressure," *Acta Otolaryngol.*, Suppl. 45, 1988, pp. 1–35.
19. E. G. Wever and M. Lawrence, *Physiological Acoustics*, Princeton University Press, Princeton, NJ, 1954.

20. P. Dallos, *The Auditory Periphery: Biophysics and Physiology*, Academic, New York, 1973.
21. J. J. Zwislocki, "The Role of the External and Middle Ear in Sound Transmission," in D. B. Tower and E. E. Eagles (Eds.), *The Nervous System, Vol. 3, Human Communication and Its Disorders*, Raven, New York, 1975, pp. 45–56.
22. W. F. Decraemer, S. M. Khanna, and W. R. J. Funnell, "Malleus Vibration Mode Changes with Frequency," *Hear. Res.*, Vol. 54, 1991, pp. 305–318.
23. W. F. Decraemer, S. M. Khanna, and W. R. J. Funnell, "A Method for Determining Three-Dimensional Vibration in the Ear," *Hear. Res.*, Vol. 77, 1994, pp. 19–37.
24. R. L. Goode, M. Killion, K. Nakamura, and S. Nishihara, "New Knowledge about the Function of the Human Ear: Development of an Improved Model," *Am. J. Otol.*, Vol. 15, 1994, pp. 145–154.
25. J. J. Zwislocki, "Analysis of Some Auditory Characteristics," in R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Wiley, New York, 1965.
26. A. Dancer and R. Franke, "Intracochlear Sound Pressure Measurements in Guinea Pigs," *Hear. Res.*, Vol. 2, 1980, pp. 191–205.
27. T. J. Lynch, III, V. Nedzelnitsky, and W. T. Peake, "Input Impedance of the Cochlea in Cat," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 108–130.
28. E. Borg, "A Quantitative Study of the Effect of the Acoustic Stapedius Reflex on Sound Transmission through the Middle Ear of Man," *Acta Otolaryngol.*, Vol. 66, 1968, pp. 461–472.
29. A. R. Møller, "An Experimental Study of the Acoustic Impedance of the Middle Ear and Its Transmission Properties," *Acta Otolaryngol.*, Vol. 60, 1965, pp. 129–149.
30. W. J. Doyle and J. T. Serok, "Middle Ear Gas Exchange in Rhesus Monkeys," *Ann. Otol. Rhinol. Laryngol.*, Vol. 103, 1994, pp. 636–645.
31. R. J. Nozza, C. D. Bluestone, D. Kardatzke, and R. Bachman, "Identification of Middle-Ear Effusion by Aural Acoustic Admittance and Otoscopy," *Ear Hearing*, Vol. 15, 1994, pp. 310–323.
32. M. Von Unge, W. F. Decraemer, J. J. Dirckx, and D. Bägger-Sjobäck, "Shape and Displacement Patterns of the Gerbil Tympanic Membrane in Experimental Otitis Media with Effusion," *Hear. Res.*, Vol. 82, 1995, pp. 184–196.
33. J. J. Zwislocki and A. Feldman, "Acoustic Impedance of Pathological Ears," ASHA Monographs Number 15, American Speech and Hearing Association, Washington, D.C., 1970.
34. J. B. Booth (Ed.), *Otology, Vol. 3*, in *Scott-Brown's Otolaryngology*, 5th ed., Butterworths, London, 1987.
35. J. B. Nadol, Jr., and H. F. Schucknecht (Eds.), *Surgery of the Ear and Temporal Bone*, Raven, New York, 1993.
36. R. T. Sataloff and J. Sataloff, *Hearing Loss*, 3rd ed., M. Dekker, New York, 1993.
37. D. T. Kemp, "Stimulated Acoustic Emissions from within the Human Auditory System," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 1386–1391.
38. P. Zurek, "Acoustic Emissions from the Ear: A Summary of Results from Humans and Animals," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 340–344.
39. B. L. Lonsbury-Martin, F. P. Harris, B. B. Stagner, B. A. Hawkins, and G. K. Martin, "Distortion-Product Emissions in Humans I. Basic Properties in Normally-Hearing Subjects," *Annals ORL*, Vol. 99, 1990, pp. 3–14.
40. B. P. Kimberley, I. Hernadi, A. M. Lee, and D. K. Brown, "Predicting Pure Tone Thresholds in Normal and Hearing Impaired Ears with Distortion Product Emissions and Age," *Ear Hearing*, Vol. 15, 1994, pp. 199–209.
41. V. Nedzelnitsky, "Sound Pressure in the Basal Turn of the Cat Cochlea," *J. Acoust. Soc. Am.*, Vol. 68, 1980, pp. 1676–1689.
42. T. J. F. Buunen and M. S. M. G. Vlaming, "Laser-Doppler Velocity Meter Applied to Tympanic Membrane Vibrations in Cat," *J. Acoust. Soc. Am.*, Vol. 69, 1980, pp. 744–750.
43. E. Zwicker and E. Schloth, "Interrelation of Different Oto-Acoustic Emissions," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 1148–1154.
44. M. E. Ravicz, J. J. Rosowski, and H. F. Voigt, "Sound-Power Collection by the Auditory Periphery of the Mongolian Gerbil *Meriones unguiculatus*: I. Middle-Ear Input Impedance," *J. Acoust. Soc. Am.*, Vol. 92, 1992, pp. 157–177.
45. J. J. Zwislocki, "An Ear-like Coupler for Earphone Calibration," Special Report LSC-S-9 of Laboratory of Sensory Communication, 1971.
46. W. M. Rabinowitz, "Measurement of the Acoustic Input Immitance of the Human Ear," *J. Acoust. Soc. Am.*, Vol. 70, 1980, pp. 1025–1035.
47. J. J. Rosowski, P. J. Davis, S. N. Merchant, K. M. Donahue, and M. D. Coltrera, "Cadaver Middle Ears as Models for Living Ears: Comparisons of Middle-Ear Input Impedance," *Ann. Otol. Rhinol. Laryngol.*, Vol. 99, 1991, pp. 403–412.
48. J. B. Allen, "Measurements of Eardrum Acoustic Impedance," in J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Eds.), *Peripheral Auditory Mechanisms*, Springer-Verlag, New York, 1986, pp. 44–51.
49. T. J. Lynch, III, W. T. Peake, and J. J. Rosowski, "Measurements of the Acoustic Input Impedance of Cat Ears: 10 Hz to 20 kHz," *J. Acoust. Soc. Am.*, Vol. 96, 1994, pp. 2184–2209.
50. H. Hudde, "Measurement of the Eardrum Impedance of Human Ears," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 242–247.
51. D. H. Keefe, J. C. Bulen, K. H. Arehart, and E. M. Burns, "Ear-Canal Impedance and Reflection Coefficient

in Human Infants and Adults," *J. Acoust. Soc. Am.*, Vol. 94, 1993, pp. 2617–2638.

52. S. E. Voss and J. B. Allen, "Measurements of Acoustic Impedance and Reflectance in the Human Ear Canal," *J. Acoust. Soc. Am.*, Vol. 95, 1994, pp. 372–384.
53. J. Jerger, S. Jerger, and L. Mauldin, "Studies in Impedance Audiometry. I. Normal and Sensorineural Ears," *Arch. Otolaryngol.*, Vol. 96, 1972, pp. 513–523.
54. A. S. Feldman and L. A. Wilber, *Acoustic Impedance & Admittance—The Measurement of Middle Ear Function*, Williams & Wilkens, Baltimore, 1976.
55. R. H. Margolis and J. E. Shanks, "Tympanometry," in J. Katz (Ed.), *Handbook of Clinical Audiology*, 3rd ed., Williams and Wilkens, Baltimore, 1985. Chap. 23.
56. Working Group on Aural Acoustic-immitance Measurements, Committee on Audiologic Evaluation, "Tympanometry," *J. Speech Hear. Disord.*, Vol. 53, 1988, pp. 354–377.
57. R. H. Margolis and J. E. Shanks, "Tympanometry: Basic Principles and Clinical Applications," in William F. Rintelmann (Ed.), *Hearing Assessment*, 2nd ed., Pro-Ed, Austin, TX, 1991, pp. 179–245.
58. J. W. Hall, III and D. Chandler, "Tympanometry in Clinical Audiology," in J. Katz (Ed.), *Handbook of Clinical Audiology*, Williams and Wilkens, Baltimore, 1994, Chap. 20.
59. J. J. Rosowski, L. H. Carney, T. J. Lynch, III, and W. T. Peake, "The Effectiveness of the External and Middle Ears in Coupling Acoustic Power into the Cochlea," in J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Eds.), *Peripheral Auditory Mechanisms*, Springer-Verlag, New York, 1986, pp. 3–12.
60. H. Rasmussen, "Studies on the Effect on the Air Conduction and Bone Conduction from Changes in Meatal Pressure in Normal Subjects and Otosclerotic Patients," *Acta Otolaryngolog.* Suppl. 74, 1948, pp. 54–64,
61. E. H. Huizing, "Bone Conduction—The Influence of the Middle Ear," *Acta Otolaryngol.* Suppl. 155, 1960, pp. 1–99.
62. H. Wada and T. Kobayashi, "Dynamical Behavior of Middle Ear: Theoretical Study Corresponding to Measurement Results Obtained by a Newly Developed Measuring Apparatus," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 237–245.
63. S. L. Naeve, R. H. Margolis, S. C. Levine, and E. M. Fournier, "Effect of Ear-Canal Air Pressure on Evoked Oto-Acoustic Emissions," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 2091–2095.
64. S. Silman (Ed.), *The Acoustic Reflex: Basic Principles and Clinical Applications*, Academic Press, New York, 1984.
65. E. Teig, "Differential Effect of Graded Contraction of Middle Ear Muscles on the Sound Transmission of the Ear," *Acta Physiol. Scand.*, Vol. 88, 1973, pp. 382–391.
66. A. L. Nuttall, "Tympanic Muscle Effects on Middle-Ear Transfer Characteristic," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1239–1247.
67. E. Borg and J-E. Zakrisson, "Stapedius Reflex and Speech Features," *J. Acoust. Soc. Am.*, Vol. 54, 1973, pp. 525–527.
68. F. B. Simmons, "Middle Ear Muscle Protection from the Acoustic Trauma of Loud Continuous Sound," *Ann. Otol. Rhinol. Laryngol.*, Vol. 69, 1960, pp. 1063–1072.
69. E. Borg and J-E. Zakrisson, "Stapedius Reflex and Monaural Masking," *Acta Otolaryngolog.*, Vol. 78, 1974, pp. 155–161.
70. X. D. Pang, "Effects of Stapedius-Muscle Contractions on Masking of Tone Responses in the Auditory Nerve," RLE Tech. Rep. No. 544, Research Laboratory of Electronics, M.I.T. Cambridge, MA, 1989.
71. R. H. Wilson and R. H. Margolis, "Acoustic-Reflex Measurement," in William F. Rintelmann (Ed.), *Hearing Assessment*, 2nd ed., Pro-Ed Inc., Austin, TX, 1991, pp. 247–319.
72. Y. Onchi, "Mechanism of the Middle Ear," *J. Acoust. Soc. Am.*, Vol. 33, 1961, pp. 794–805.
73. A. R. Møller, "Network Model of the Middle Ear," *J. Acoust. Soc. Am.*, Vol. 33, 1961, pp. 168–176.
74. J. Zwislocki, "Analysis of the Middle-Ear Function. Part I: Input Impedance," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 1514–1523.
75. J. W. Matthews, "Modeling Reverse Middle Ear Transmission of Acoustic Distortion Signals," in E. de Boer and M. A. Viergever (Eds.), *Mechanics of Hearing*, Martinus Nijhoff, The Hague, 1983, pp. 11–17.
76. M. Kringlebotn, "Network Model for the Human Middle Ear," *Scand. Audiol.*, Vol. 17, 1988, pp. 75–85.
77. W. M. Rabinowitz, "Acoustic-Reflex Effects on the Input Admittance and Transfer Characteristics of the Human Middle Ear," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1977.
78. W. R. J. Funnell, S. M. Khanna, and W. F. Decraemer, "On the Rigidity of the Manubrium in a Finite Element Model of the Cat Eardrum," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 2081–2090.
79. W. T. Peake, J. J. Rosowski, and T. J. Lynch, III, "Middle-Ear Transmission: Acoustic versus Ossicular Coupling in Cat and Human," *Hear. Res.*, Vol. 57, 1992, pp. 245–268.
80. C. A. Shera and G. Zweig, "Middle-Ear Phenomenology: The View from the Three Windows," *J. Acoust. Soc. Am.*, Vol. 92, 1992, pp. 1356–1369.
81. M. Kringlebotn and T. Gundersen, "Frequency Characteristics of the Middle Ear," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 159–164.
82. M. Kringlebotn, "The Equality of Volume Displacements in the Inner Ear Windows," *J. Acoust. Soc. Am.*, Vol. 98, 1995, pp. 192–196.
83. J. L. Goldstein, "Auditory Nonlinearity," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 676–689.

107

ANATOMY OF THE COCHLEA AND AUDITORY NERVE

NORMA B. SLEPECKY

1 INTRODUCTION

The cochlea contains the sensory cells, supporting cells, and nerve fibers of the inner ear that are responsible for detection of sound. When sound pressure waves vibrate the tympanic membrane, the movement is conducted through the middle-ear space by the middle-ear bones (the malleus, incus, and stapes). The stapes passes these vibrations into the fluid-filled cochlea. The vibrations are distributed in a frequency-specific manner along the length of the sensory organ to stimulate individual hair cells and nerve fibers in the sensory organ of Corti.

It was thought initially that the organ of Corti was purely a passive detector of mechanical vibration and that normal frequency selectivity depended largely on the structural gradients of the component parts. However, it is now known that the hair cells resond to stimulation, and this response actively increases the threshold sensitivity and frequency selectivity of this exquisitely designed mechanoreceptive organ.

In order to understand the crucial steps in the process by which mechanical vibrations are transduced into electrical responses and neural impulses, it is necessary to understand the anatomy and cell biology of the sensory organ involved. The goal of this chapter is to introduce the cochlea—the sensory and supporting cells, the nerve fibers, and the cells of the stria vascularis and spiral ligament that are responsible for the production of endolymph and the maintenance of the endocochlear potential.

2 THE COCHLEA

The cochlea is actually a cavity in the temporal bone having the shape of a coiled tube. The outside of the spiral cochlea is defined by the otic capsule (Fig. 1). Within the cochlea, the membranous labyrinth contains the cells, structures, and fluids necessary for the detection of sound. The membranous labyrinth is subdivided into three anatomically defined compartments or scalae: scala vestibuli, scala media, and scala tympani (Fig. 2*a*). These three compartments separate two fluids; perilymph, which is high in sodium and typical of the type of fluids found outside of cells elsewhere in the body, and endolymph, which is high in potassium similar in composition to fluids found inside of cells. It is commonly stated that perilymph is contained in both scala vestibuli and scala tympani while endolymph is found only within the region of scala media (Fig. 2*a*). However, the actual boundaries of the fluid-filled compartments are not the same as the anatomically defined compartments described above. The endolymph containing space is in fact limited by the tight junctions between the adjacent cells of Reissner's membrane, the top surfaces of the sensory and supporting cells of the organ of Corti (forming the reticular lamina), and the marginal cells of the stria vascularis where the blood vessels run along the lateral wall (Fig. 2*b*). Thus, the sensory and supporting cells of the receptor organ of Corti have endolymph on one surface and a perilymphlike fluid on the other. The nerve terminals and fibers are bathed in a perilymphlike fluid (sometimes referred to as Cortilymph) that is high in sodium.

3 THE ORGAN OF CORTI

In the organ of Corti, the sensory and supporting cells rest on the basilar membrane. The sensory organ contains (depending on the species studied) approximately 20,000 sensory hair cells, which are in contact with various types of supporting cells. The sensory cells are described as *inner* hair cells, which form one row along the length of

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

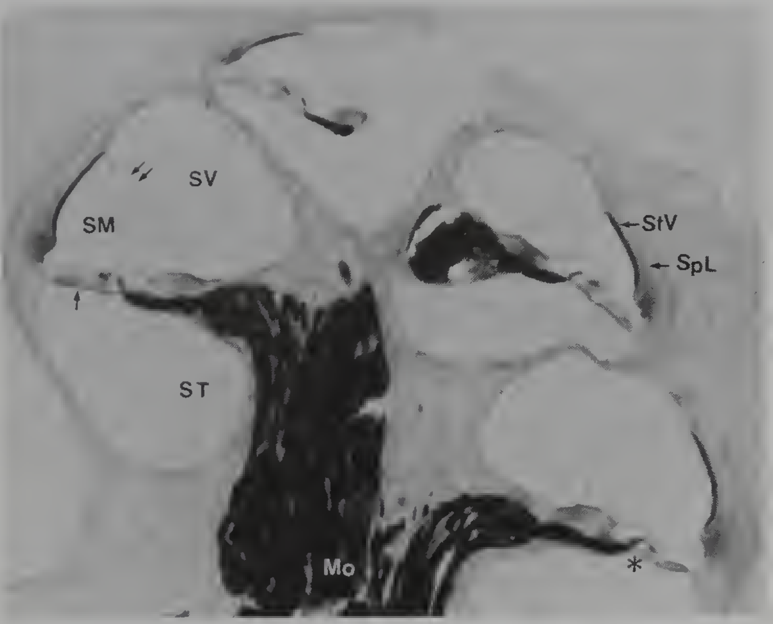
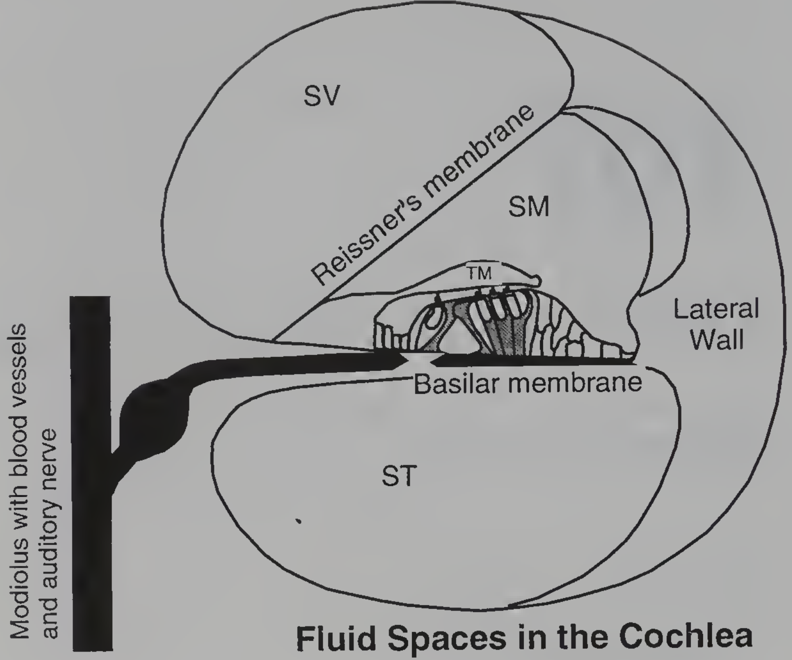


Fig. 1 Midmodiolar section of the rabbit cochlea, showing the membranous labyrinth encased in the bone of the otic capsule. Reissner's membrane (double arrow) and the basilar membrane (single arrow) form the boundaries of scala media (SM) and separate it from scala vestibuli (SV) and scala tympani (ST). The stria vascularis (StV) and the spiral ligament (SpL) lie along the lateral wall. The organ of Corti (*) is located on the basilar membrane and the nerve fibers supplying it run in the modiolus (Mo).

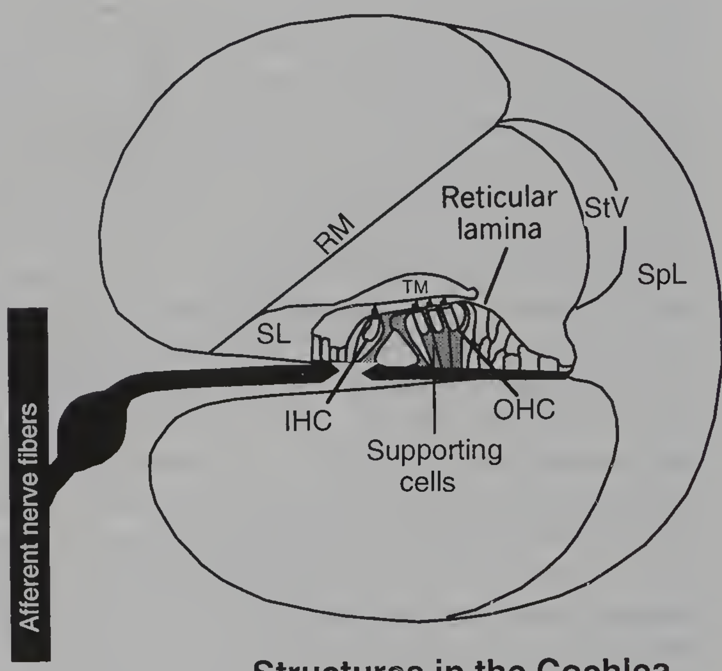
the sensory epithelium closer to the central axis of the spiral and the modiolus in which the auditory nerve fibers are found, and *outer* hair cells, which are found in three rows and which are separated from the inner hair cells by the inner and outer pillar supporting cells (Fig. 2*b*). During stimulation by sound, movement of the basilar membrane is transmitted through the supporting cells so that the top surfaces of the sensory and supporting cells vibrate at the same frequency as the stimulus. Shearing of the stereocilia, the hairs on the sensory hair cells, against the gelatinous mass of the overlying tectorial membrane results in movement of the stereocilia and is the event by which mechanical movement is transduced into an electrical response in the sensory cells.

The organ of Corti appears to be loosely structured because of the tremendous amount of fluid-filled spaces surrounding the sensory and supporting cells. However, cell structure and organization appears to function optimally in the transmission of movement of the basilar membrane to the apical surface of the sensory hair cells in response to both low- and high-frequency sound. The sensory cells are designed for maximal sensitivity to small motions, and the supporting cells are structured for rigid support. The cellular and structural arrangement of cells within the organ of Corti have been well studied in many different species of animals, and for the most part they appear similar.[1–3] In most species, the size of the organ of Corti changes continuously along its length from apex to base. In the apex of the cochlea, cells are usually larger, stereocilia are longer and less stiff, the basilar membrane is wider and less stiff, and the tecto-



Fluid Spaces in the Cochlea

(*a*)



Structures in the Cochlea

(*b*)

Fig. 2 (*a*) Boundaries of the endolymph containing compartment called scala media (SM) are Reissner's membrane, the basilar membrane, and the lateral wall. Both scala vestibuli (SV) and scala tympani (ST) contain perilymph. (*b*) Reticular lamina of the organ of Corti is made up of the apical surfaces of the inner (IHC) and outer (OHC) hair cells and the supporting cells. The tectorial membrane (TM) projects out from the spiral limbus (SL) and overlies the organ of Corti. The stria vascularis (StV) and the spiral ligament (SpL) line the lateral wall. Afferent nerve fibers from IHC and OHC run in the modiolus.

rial membrane has a greater mass than at the base. The absolute size varies from species to species and is probably related to the tuning capacities and sensitivities of the receptor organs. In some species there are regions with exaggerated structural features that appear only in localized regions and are correlated with specialized functions.

3.1 Sensory Hair Cells

The sensory hair cells are specialized to detect mechanical stimulation and to transmit information to nerve fibers that connect with the central nervous system. On their apical surface, the sensory hair cells have stereocilia,[4] which appear as stiff extensions present in rows arranged in a U or W shape. The stereocilia are graduated in length (Figs. 3 and 4), with the shortest in a row on the side of the hair cell nearer the central region of the cochlear spiral and the tallest in a row on the side nearer the lateral wall. All the stereocilia on one hair cell are linked together.[5] These links include side-to-side links that connect stereocilia within one row and tip links that connect the top of one stereocilium with its taller neighbor in the adjacent row (Fig. 5). Since actin filaments are present in the stereocilia and crosslinked by actin-binding proteins that form the filaments into rigid bundles, it is thought that the stereocilia are stiff and pivot at their base where they are inserted into the cuticular plate. Deflection of the stereocilia in one direction opens ion-selective channels at the tips of the stereocilia; potassium ions flow into the cell, the cell becomes depolarized, and neurotransmitter is released from the cell at an increased rate.

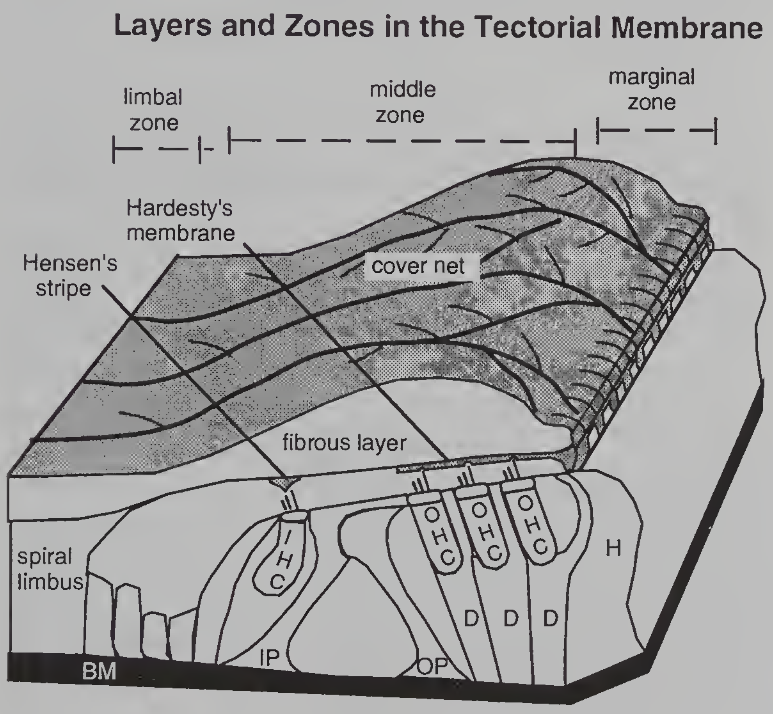


Fig. 3 Organ of Corti rests on the basilar membrane (BM). In the organ of Corti there are both IHC and OHC sensory hair cells. The inner (IP) and outer (OP) pillar cells form the tunnel of Corti. The Deiters (D) cells support the OHCs. The stereocilia from the IHC and OHC project into the tectorial membrane, which can be subdivided into several zones and layers. The limbal zone covers the spiral limbus. The middle zone overlies the sensory cells of the organ of Corti. The marginal zone is at the lateral edge of the TM. The cover net forms the superficial layer of the TM and becomes contiguous with Hardesty's membrane in the region of the marginal zone. The body of the TM is comprised mainly of the fibrous layer. Hensen's stripe and Hardesty's membrane are structures of the TM that form continuous regions (from the base of the cochlea to the apex, that are most closely associated with the tips of the stereocilia of the sensory hair cells.

Based on careful analysis of sensory hair cells, supporting cells and nerve fibers, it is now clear that inner hair cells (IHCs) differ significantly from outer hair cells (OHCs). Inner hair cells are goblet shaped with a centrally placed nucleus. Their cytoplasm contains a network of filaments and tubules that act to support the cell and maintain the cell shape. Each cell contains all of the machinery for protein synthesis. The cells have stereocilia at their top surface, which are anchored into a filament network at one end within the cell and that project into the endolymphatic fluid at their other end. The sides of the cells are in contact with supporting cells, which span the distance between the basilar membrane and the reticular lamina. The bottom ends of the cells make direct contact with the nerve fibers projecting to the central nervous system. Because the IHC afferent nerve fibers comprise 95% of the fibers within the auditory nerve, it is thought that IHCs are the primary receptor cells in the cochlea. Efferent nerve fibers carrying information from the brain to the organ of Corti make direct contact with the afferent nerve fibers in this region, but not with the IHCs.

Outer hair cells are long and cylindrical in shape, with a more basally placed nucleus (Fig. 5). They form three rows along the length of the organ of Corti, they are slightly smaller in diameter than IHCs, and they number almost four times as many. At their top surface, the OHCs contact supporting cells and at their bottom surfaces they are surrounded by supporting cells. Unlike IHCs, they contact only fluid along their lateral surfaces. Stereocilia project out from the top surface of the OHCs, and the tips of the stereocilia from the tallest row are embedded in the tectorial membrane. Above the nucleus, there is some evidence for the machinery required for protein synthesis. Along the sides of the cylinder, there is a prominent layer of membranous sacs lined with mitochondria that provide energy to the cell. At the bottom of the OHCs, the cells make direct contact with both affer-

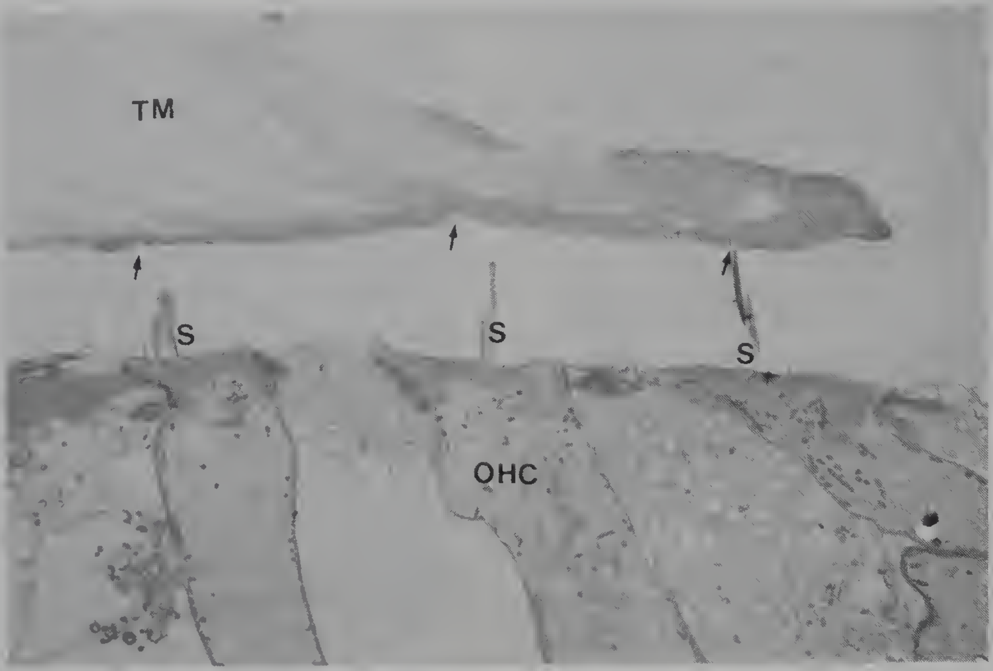


Fig. 4 In a higher magnification micrograph, it can be seen that there are indentations in the bottom surface (arrows) of the tectorial membrane (TM) where the tips of the stereocilia (S) of the outer hair cells (OHC) are embedded.

ent nerve fibers going to the brain and efferent nerve fibers carrying information from the brain back out to the organ of Corti.

During stimulation by sound, OHCs (but not IHCs) shorten and elongate at both a slow and a fast rate. It has been suggested that these motile responses result in active processes that have been described in the cochlea, including the establishment of sharp frequency tuning and the generation of otoacoustic emissions. Two distinct structural differences between OHCs and IHCs could be responsible for their functional differences in the cochlea. First, efferent nerve fibers directly contact OHCs while they synapse only on afferent nerve fibers below the IHCs. Second, the OHCs appear to have a specialized cell membrane that is responsible for the fast OHC shortening and elongation.

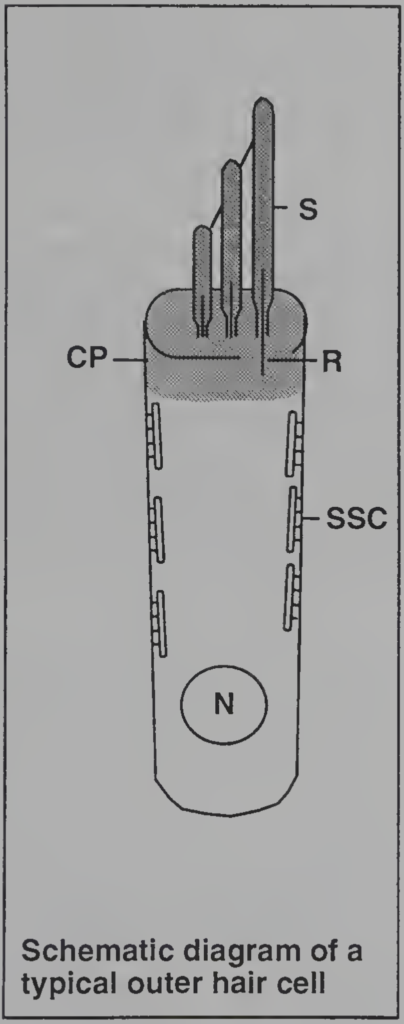


Fig. 5 Schematic diagram of a typical outer hair cell. Along the apical surface, stereocilia (S) extend out from the cuticular plate (CP). At its base, each stereocilium has a rootlet (R) that projects into the cuticular plate, and a tip link that connects it to its taller neighbor. Along the lateral wall there are membranous sacs, or subsurface cisternae (SSC). The nucleus (N) is located near the base of the cell.

3.2 Supporting Cells

Supporting cells in the organ of Corti have not been described in as great detail as the sensory cells; however, they have critical roles in providing structural and metabolic support. Based on their anatomy as seen at the light microscope level, they can be divided into two groups—cells with filaments and tubules in bundles forming a rigid scaffolding (to provide physical support) and cells with less organized skeletal components yet containing organelles necessary for protein synthesis and ion transport (to maintain a normal balance of ions in the cells and fluids).

The inner tunnel pillar cells, outer tunnel pillar cells, and Deiters cells are structurally similar and contain large numbers of filaments and tubules in bundles.[6] These cells are found in the region of the organ of Corti where it rests on the basilar membrane and is not directly supported by any bone. This region is critical for stimulation of the hair cells, since vibration of the basilar membrane is transmitted through the supporting cells to the stereocilia, and movement of the stereocilia results in transduction of mechanical stimulus into an electrical response in the sensory hair cells. Thus, it is not surprising that these supporting cells are specialized and contain microfilaments and microtubules to reinforce them and allow them to better withstand mechanical stress. Unlike the sensory cells, these supporting cells also contain an additional type of filaments, intermediate in size between the microtubles and microfilaments, which play a role in the maintenance of stable cell shape.[7]

Less is known about the supporting cells, which lack organized filament and tubule bundles. Hensen cells are directly adjacent to the third row of Deiters cells. In the apical turns of the cochlea, they reach from the reticular lamina to the basilar membrane. In the basal turns of the cochlea they rest on Boettcher cells. While they contain no filament or tubule bundles, they may be distinguished from the other cells in this region by their prominent oil droplets (the function of which is not yet known). Other cells that are part of the supportive system include Claudius cells, which extend from the organ of Corti to the lateral wall of the cochlea. External sulcus cells, or root cells, are located at the junction of the basilar membrane with the cells of the lateral wall. All of these supporting cells appear similar to each other morphologically,[8] and it is believed that they participate in the formation of the fluid surrounding the hair cells rather than providing structural support.

4 ASSOCIATED STRUCTURES

4.1 Tectorial Membrane

Although the exact mechanism that initiates sensory cell excitation in the organ of Corti is unknown, it is generally accepted that moving the stereocilia on the apical surface of the sensory hair cells by a shearing movement of the tectorial membrane (TM) causes sensory cell excitation. The structure and composition of the TM is of particular interest because it is in a unique position to affect the transduction process. The TM is anchored at the spiral limbus (Fig. 6), and overlies the stereocilia of the sensory hair cells. Contrary to the early anatomical findings, which suggested that the TM was an amorphous mass with no internal organization, it is now known that the TM is composed of at least two distinct types of fibers.[9] Type A fibers are straight, unbranched, and 10 nm in diameter, running radially in parallel bundles within the fibrous layer of the TM. Type B fibers

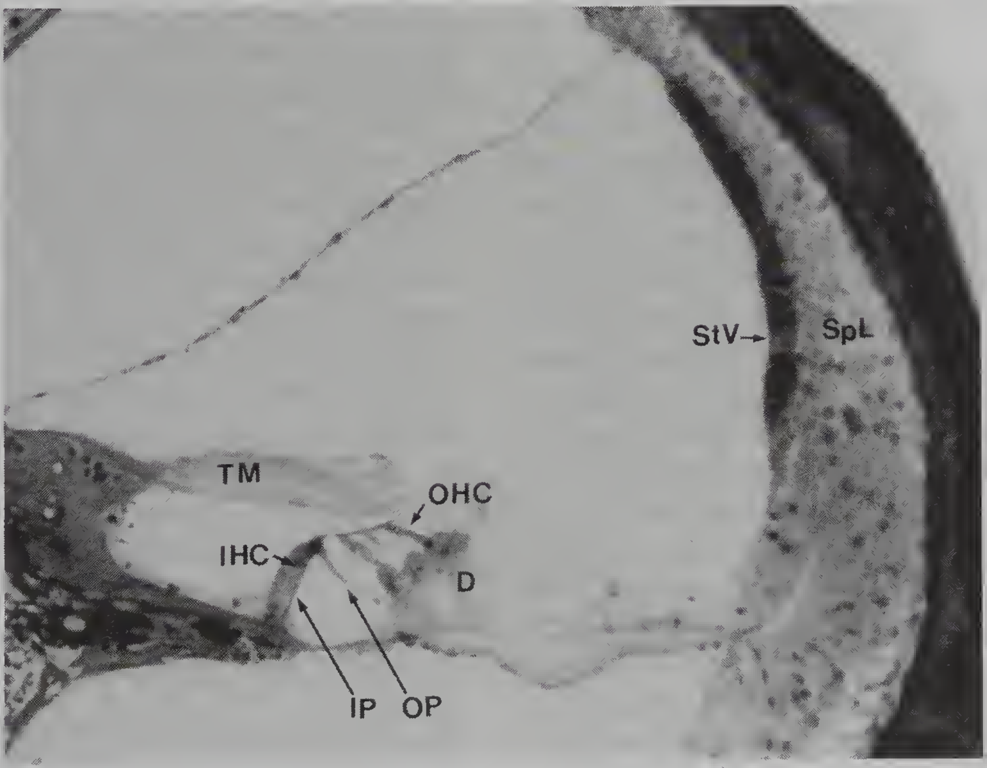


Fig. 6 In a cross section through one turn of the cochlea, the tectorial membrane (TM), inner hair cell (IHC), outer hair cells (OHCs), and supporting cells (IP, OP, and D) of the organ of Corti can be seen even at the light microscopic level. Along the lateral wall, the stria vascularis (StV) appears more darkly stained than does the spiral ligament (SpL).

are tightly coiled 4-nm filaments that are wavy, highly branched, and interconnected.[10]

The distribution of the type A and type B fibers and the different packing arrangement of type B fibers divide the TM into three layers (running parallel to the reticular lamina) and three zones[11] (Fig. 3). The regions over the sensory hair cells are specialized for contact with the stereocilia. A structure called Hardesty's membrane forms a continuous layer over the region of the OHCs, and there are indentations where the tips of the longest stereocilia insert into it. Hensen's stripe is the thickened region that overlies the IHCs, but consistent demonstration of direct contact of the tips of the IHC stereocilia with the TM is still lacking.

It is now clear that the TM contains at least three types of the protein collagen.[12–14] The type A fibers are composed of type II and type IX collagens in bundles that are embedded in a network of type B fibers that contain type V collagen. Thus, the TM appears to be composed of materials (collagen fibers and other proteins associated with carbohydrates that surround the collagen fibers) that are similar in composition to those found in cartilage. Type II collagen fibrils are thought to contribute to the structural stabilization and tensile strength of the TM and provide support as the TM is subjected to stress during vibration of the organ of Corti. The other components can influence the amount of water within the fiber matrix of the TM and influence the amount of crosslinking of the component materials. Thus, the components of the TM will determine not only the structure but also the biomechanical properties of the TM.

4.2 Basilar Membrane

The organ of Corti rests on a structure called the basilar membrane (BM), which contributes to the stiffness and mass of the cochlear partition. In normal adult animals, the stiffness increases progressively toward the base while the mass decreases, and this gradient provides the physical basis for the largely passive mechanical tuning seen in the cochlea. In the apex of the cochlea it is wide (spanning a large distance between the osseous spiral lamina and the spiral ligament at the lateral wall) and thin while at the base of the cochlea it becomes narrow (spanning a short distance between the osseous spiral lamina and the spiral ligament at the lateral wall) and thick. It is composed mainly of bundles and networks of fibers that are produced by cells in the area.[15] In this region there are several proteins present including type II collagen, type IX collagen[16] and fibronectin.[17] The region called the pars tecta (arcuate zone) extends from the spiral limbus to the region under the outer pillar cells. The region called the pars pectinata (pectinate zone) extends from the region under the outer pillar cells to the spiral ligament along the lateral wall. The side of the BM facing scala tympani is covered with tympanic border cells. Because there are large spaces between adjacent cells in this region, it is thought that the BM is in direct contact with perilymph and that perilymph permeates through the BM to contribute to the fluid surrounding the sides and bottom surfaces of the sensory and supporting cells.

5 BLOOD SUPPLY

The blood supply to the cochlea is compartmentalized into at least three large and spatially separate capillary beds. Extensive studies[18] have determined that there are similarities in the blood supply to the cochleas in different mammalian species. Large arteries and veins spiral in the modiolus. Small arteries (arterioles) branch off to form capillary beds in the region of the cochlea between the modiolus and the organ of Corti. Some arterioles radiate out through the bony partition of each turn, over the scala vestibuli to form distinct capillary beds in the spiral ligament and stria vascularis. The capillaries are drained by small veins (venules) that empty into one or two larger veins running spirally around the modiolus. The capillaries in the stria vascularis form the most dense network in the lateral wall. Cells making up the walls of the capillaries are tightly coupled so that substances in the blood cannot pass out into the cochlear tissue between adjacent cells. It has been noted that the capillaries form a barrier between the blood and the strial cells that is more restrictive than the barrier between the blood and the neurons of the brain.

It is surprising, with respect to the metabolism of the sensory and supporting cells, that the area near the organ of Corti is essentially devoid of capillaries. Thus, metabolites and oxygen must diffuse in the fluids to reach the sensory cells—a process that might not be rapid enough to keep up with increased demand during noise trauma. The regions of the cochlea that do have dense capillary networks also have cells that are specialized for producing fluids. Based on the cell types present, the proximity to capillaries and fluid-filled spaces of the cochlea, it is thought that these regions are either involved in maintaining and/or producing perilymph.

6 LATERAL WALL

The stria vascularis and the spiral ligament are the cells along the lateral wall of the cochlea that are thought to play a critical role in the production of cochlear fluids (maintaining the high concentration of potassium in endolymph) and maintaining the endolymphatic potential in scala media. Both of these functional roles are required for optimal transduction and our perception of sound.

6.1 Stria Vascularis

The stria vascularis is the region along the lateral wall of the cochlea where there is a high density of capillaries and morphological specializations characteristic of fluid transporting cells. Three cell types are present in the stria vascularis (Fig. 7): marginal cells, intermediate cells, and basal cells. The marginal cells and the basal cells are tightly coupled along their lateral surfaces so that they form a separate fluid-filled compartment. Marginal cells line the endolymphatic surface of the scala media. These cells are specialized for high rates of metabolism and transport of fluids and ions.[19,20] The intermediate cells form a discontinuous layer beneath the marginal cells. They do not contact the endolymph, and their intermediate distribution between the marginal and basal cells allows them to bridge the other two cell types and play a role in generating the positive component of the endocochlear potential. These cells do not show evidence of high rates of metabolism but are capable of synthesizing melanin pigment granules.[21] In mutants where intermediate cells are lacking, the endocochlear potential is close to zero suggesting that intermediate cells are necessary either for the development or activity of marginal cells or for providing potassium for movement into endolymph.[22] The basal cells are flat, overlapping cells that form a continuous cell layer separating the marginal cells, intermediate cells, and capillaries of the stria vascularis from the spiral ligament and perilymph.

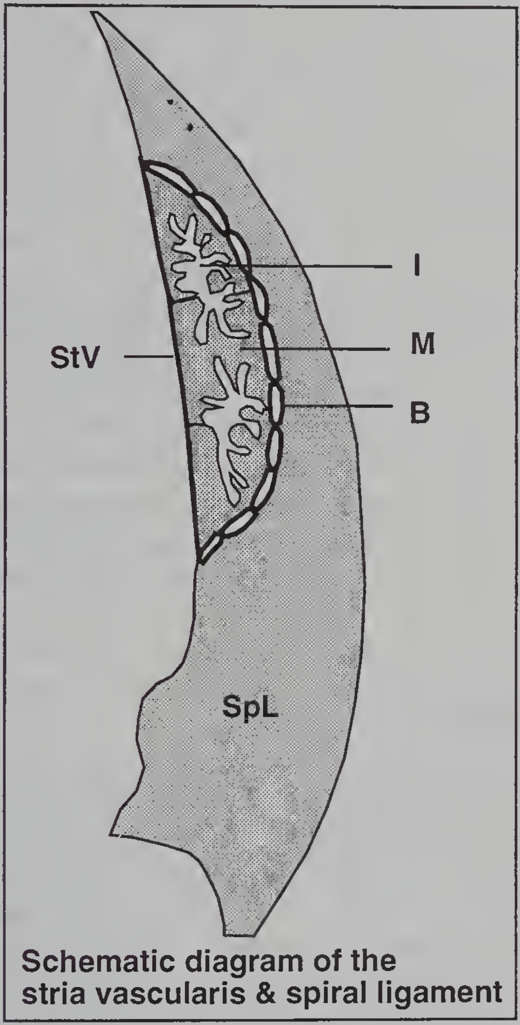


Fig. 7 Schematic diagram of the stria vascularis and spiral ligament along the lateral wall of the cochlea. The strial vascularis (StV) is composed of marginal cells (M), intermediate cells (I), and basal cells (B) as well the major capillary bed. The spiral ligament (SpL) contains much extracellular fibrous material and specialized ion transporting fibroblasts.

6.2 Spiral Ligament and Spiral Prominence

The spiral ligament is a region of the cochlea adjacent to the otic capsule (Figs. 1 and 7), where there are only a small number of cells embedded in a network of fibers. It provides support for the capillaries in the stria vascularis and may absorb stresses that are generated by vibration of the BM, with which it is structurally linked. It can be divided into five regions based on the type and shape of cells present[23]—some of which are morphologically and biochemically specialized for transporting fluids and ions.[19,20] The spiral prominence is a region that contains a large number of capillaries and it bulges into scala media near the point at which the BM inserts.

7 INNERVATION

Early studies in the cochlea showed that there was more than one type of nerve fiber running between the organ of Corti and the central nervous system. One set of nerves contacting the sensory cells of the cochlea link them directly to the central nervous system. Neurons projecting from the sensory hair cells in each cochlea go only to the cochlear nuclear complex on the same (ipsilateral) side of the brain. It is here that nerve fibers branch, and there is clear evidence for the establishment of parallel processing of the primary sensory information carried in the auditory nerve. Another set of nerve fibers project from the central nervous system out to the organ of Corti, where they are in a position to influence the responses of the auditory nerve fibers and the OHCs directly. Anatomical studies were instrumental in identifying the fibers as afferent or efferent and in determining the number of afferent and efferent nerve endings contacting each type of hair cell in each turn of the cochlea. Since then, many reports on the innervation pattern, structure of the contact site between the hair cell and the nerve fiber, and the chemical that signals the transmission of information have appeared.[24]

7.1 Afferents

The point of contact where information is passed from one cell to another is called the synapse. The synapse between the hair cell and the afferent nerve fiber is characterized by a localized thickening of the hair cell membrane and the presence in the hair cell of small vesicles thought to contain the chemical that will be released from the cell to signal the adjacent nerve fiber. The synapse between an efferent nerve fiber and a hair cell is characterized by accumulations in the nerve fiber terminal of small vesicles. In regions of the cochlea away from the hair cells, the two populations of nerve fibers are difficult to distinguish because the bundles contain both types of nerve fibers. Thus, analysis has been accomplished by techniques that selectively identify fiber types (observing degeneration of the nerve fibers after cutting them in regions where the populations are physically separated, following the pathway of individual fibers in adjacent sections over long distances, injecting tracer substances at one place and seeing where they end up, use of specific markers for the different chemical transmitters used by the different types of nerve fibers).

Based on these results, it has been confirmed that the afferent fibers are subdivided into at least two systems. Early studies[25,26] showed that 90–95% of the fibers in the auditory nerve are type I fibers. Type I fibers originate from type I ganglion cells, are thick, surrounded by myelin sheaths, and contact only IHCs. They are unbranched so that each fiber terminates only once and with only one IHC. Each IHC contacts approximately 20 different afferent nerves and these afferent nerves innervate all the major regions of the cochlear nucleus.

Type II fibers originate from type II ganglion cells (5% of the population) and are thin and lack myelin sheaths. These fibers are highly branched and contact multiple OHCs. The type II ganglion cells are smaller than the type I cells, and processes from these cells have been traced to the small cap regions of the cochlear nucleus.

7.2 Efferents

Efferent fibers coming from the central nervous system[27] enter the cochlea near the basal and middle turns of the cochlea. They spiral in both apical and basal directions as the intraganglionic spiral bundle, running through the spiral ganglion cells in Rosenthal's canal. At a given point, depending on the frequency specificity of the individual cell bodies of origin, the efferent fibers turn out toward the organ of Corti in a radial direction, and run with the afferent fibers in the osseous spiral lamina. They pass through to the organ of Corti with the afferent nerve fibers, and like them are unmyelinated within the organ of Corti. Although the divisions of the efferent innervation to the cochlea are at present based on which type of hair cell the fibers contact (IHC or OHC), it must be noted that subdivisions are possible based on the presence of different chemical substances that are used as neurotransmitters.[24]

The fibers to the IHC region are thin, unmyelinated, and originate from cell bodies associated with the lateral superior olivary complex on either the same side or opposite side of the brainstem. The efferent fibers that remain in the IHC region spiral in bundles beneath the IHCs or spiral in the tunnel of Corti, where they branch and contact the afferent fibers present beneath the IHCs (not usually making contact directly with IHCs). In cat, the lateral IHC efferents are most often found synapsing with low spontaneous rate, high-threshold afferent fibers.[28]

The efferent nerve fibers contacting the OHCs are from large diameter, myelinated efferent fibers originating from cell bodies close to the medial superior olivary nuclei. Projections may be either from the same side or opposite side of the brain stem, although a small component of the OHC efferent fibers arise from neurons projecting to both cochleas. The nerve fibers enter the organ of Corti and run, without branching, near the IHCs. There they can spiral for a short distance or pass through the tunnel of Corti at a midlevel to reach the OHC region. The fibers branch almost immediately in all directions and run in bundles that spiral under the OHCs prior to contacting the OHCs directly. There are differences between the innervation of the OHCs in the different rows, with more efferent contacts on the first row of OHCs at each place along the length of organ of Corti. In most species, there is a base-to-apex gradation of efferent innervation with a larger number of efferent terminals ending on each OHC at the base of the cochlea.

8 SUMMARY

The observations of many anatomists, cell biologists, and biochemists have provided us with clues as to the roles played by the different cell types in the cochlea. It is clear that the sensory hair cells are not only passive mechanoreceptors but provide active responses to stimulation that in turn affects cochlear function. Supporting cells provide both structural and metabolic support, and cells within the stria vascularis and spiral ligament interact in complex ways to transport ions and water to maintain cochlear fluid balance and endocochlear potentials. However, cells do not act alone and hearing depends critically on interactions between them. Many of the most interesting problems in these areas of cochlear research are now focused on the nature of these interactions and their role in establishing and maintaining basic peripheral mechanisms in hearing.

REFERENCES

1. S. M. Echteler, R. R. Fay, and A. N. Popper, "Structure of the Mammalian Cochlea," in R. R. Fay and A. N. Popper (Eds.), *Comparative Hearing: Mammals*, Springer-Verlag, New York, 1994, pp. 134–171.
2. D. J. Lim, "Functional Structure of the Organ of Corti: A Review," *Hear. Res.*, Vol. 22, 1986, pp. 117–146.
3. N. B. Slepecky, "Structure of the Mammalian Cochlea," in P. Dallos, A. N. Popper, and R. R. Fay (Eds.), *The Cochlea: Springer Handbook of Auditory Research*, Springer-Verlag, New York, 1996.
4. D. W. Nielsen and N. B. Slepecky, "Stereocilia," in R. A. Altschuler, R. P. Bobbin, and D. W. Hoffman (Eds.), *Neurobiology of Hearing: The Cochlea*, Raven Press, New York, 1986, pp. 23–46.
5. J. O. Pickles, S. D. Comis, and M. P. Obsorne, "Cross-links between Stereocilia in the Guinea Pig Cochlea," *Hear. Res.*, Vol. 18, 1984, pp. 177–188.
6. C. Angelborg and H. Engstrom, "Supporting Elements in the Organ of Corti," *Acta Otolarynagol.*, Vol. Suppl. 301, 1972, pp. 49–60.
7. Y. Raphael, G. H. Marshak, A. Barash, and B. Geiger, "Modulation of Intermediate-Filament Expression in Developing Cochlear Epithelium," *Differentiation*, Vol. 35, 1987, pp. 151–162.
8. R. S. Kimura, "The Ultrastructure of the Organ of Corti," *Int. Rev. Cytol.*, Vol. 42, 1975, pp. 173–222.
9. A. Kronester-Frei, "Ultrastructure of the Different Zones of the Tectorial Membrane," *Cell Tiss. Res.*, Vol. 193, 1978, pp. 11–23.
10. J. A. Hasko and G. P. Richardson, "The Ultrastructural Organization and Properties of the Mouse Tectorial Membrane Matrix," *Hear. Res.*, Vol. 35, 1988, pp. 21–38.
11. D. J. Lim, "Fine Morphology of the Tectorial Membrane," *Arch. Otolaryngol.*, Vol. 96, 1972, pp. 199–215.
12. N. B. Slepecky, J. E. Savage, L. K. Cefaratti, and T. J. Yoo, "Electron Microscopic Localization of Type II, IX and V Collagen in the Organ of Corti," *Cell Tiss. Res.*, Vol. 267, 1992, pp. 413–418.
13. G. P. Richardson, I. J. Russell, V. C. Duance, and A. J. Bailey, "Polypeptide Composition of the Mammalian Tectorial Membrane," *Hear. Res.*, Vol. 25, 1987, pp. 45–60.
14. P. Munyer and B. Schulte, "Immunohistochemical Identification of Proteoglycans in Gelatinous Membranes of Cat and Gerbil Inner Ear," *Hear. Res.*, Vol. 52, 1991, pp. 369–378.
15. S. Iurato, *Submicroscopic Structure of the Inner Ear*, Pergamon Press, Oxford, 1967, 1.
16. I. Thalmann, "Collagen of Accessory Structures of Organ of Corti," *Connective Tiss. Res.*, Vol. 29, 1993, pp. 199–201.
17. E. M. Keithley, A. F. Ryan, and N. K. Woolf, "Fibronectin-like Immunoreactivity of the Basilar Membrane of Young and Aged Rats," *J. Comp. Neurol.*, Vol. 327, 1993, pp. 612–617.
18. A. Axelsson and A. Ryan, "Comparative Study of the Vascular Anatomy in the Mammalian Cochlea," in A. F. Jahn and J. R. Santos-Sacchi (Eds.), *Physiology of the Ear*, Raven Press, New York, 1988, pp. 295–316.
19. T. Kikuchi, R. S. Kimura, D. L. Paul, and J. C. Adams, "Gap Junction Systems in the Rat Cochlea: Immunohistochemical and Ultrastructural Analysis," *Anat. Embryol.*, Vol. 191, 1995, pp. 101–118.
20. B. A. Schulte and J. C. Adams, "Distribution of Immunoreactive Na+, K+ -ATPase in Gerbil Cochlea," *J. Histochem. Cytochem.*, Vol. 37, 1989, pp. 127–134.
21. D. A. Hilding and R. D. Ginzburg, "Pigmentation of the Stria Vascularis," *Arch. Otolaryngol.*, Vol. 84, 1977, pp. 24–37.
22. B. A. Schulte and K. P. Steel, "Expression of α and β Subunit Isoforms of Na, K-ATPase in the Mouse Inner Ear and Changes with Mutations at the Wv or Sld Loci," *Hear. Res.*, Vol. 78, 1994, pp. 259–260.
23. M. M. Henson and O. W. Henson, "Tension Fibroblasts and the Connective Tissue Matrix of the Spiral Ligament," *Hear. Res.*, Vol. 35, 1988, pp. 237–258.
24. M. Eybalin, "Neurotransmitters and Neuromodulators of the Mammalian Cochlea," *Physiol. Rev.*, Vol. 73, 1993, pp. 309–373.
25. H. Spoendlin, "The Innervation of the Cochlear Receptor," in A. Moller (Ed.), *Basic Mechanisms in Hearing*, Academic Press, New York, 1983, pp. 182–230.
26. A. Schwartz, "Auditory Nerve and Spiral Ganglion Cells; Morphology and Organization," in R. A. Altschuler, R. P. Bobbin, and D. W. Hoffman (Eds.), *Neurobiology of Hearing: The Cochlea*, Raven Press, New York, 1986, pp.271–282.
27. W. B. Warr, "Organization of Olivocochlear Efferent Systems in Mammals," in D. B. Webster, A. Popper, and R. R. Fay (Eds.), *Mammalian Auditory Pathways: Neuroanatomy*, Springer-Verlag, New York, 1992, pp. 410–448.
28. M. C. Liberman, L. W. Dodds, and S. Pierce, "Afferent and Efferent Innervations of the Cat Cochlea: Quantitative Analysis with Light and Electron Microscopy," *J. Comp. Neurol.*, Vol. 301, 1990, pp. 443–460.

108

COCHLEAR MECHANICS AND BIOPHYSICS

MARIO A. RUGGERO AND JOSEPH SANTOS-SACCHI

1 INTRODUCTION

The present chapter focuses on the function of the mammalian hearing organ, the cochlea of the inner ear (whose structure is described in Chapter 107). All groups of vertebrate animals possess hearing organs, but their morphology and physiology differ greatly across the evolutionary tree (Chapter 147; for more complete review see Ref. 1). "Cochlear mechanics and biophysics" refers to the processes whereby the inner ear encodes sound into electrical signals. Figure 1 presents a block diagram of the peripheral auditory system of mammals, with arrows indicating the direction of signal transmission. Sound arriving at the external ear (Chapter 105) sets in motion tiny bones (ossicles) in the middle ear (Chapter 106); these, in turn, create pressure fluctuations in the cochlear fluids, causing a displacement wave to propagate along the basilar membrane. The displacement wave stimulates cellular transducers, the inner and outer hair cells (IHCs and OHCs), to generate, much like tiny microphones, electrical receptor potentials that mimic the acoustic stimulus. Finally, these receptor potentials produce chemically mediated excitation in the peripheral terminals of cochlear afferent neurons, generating trains of electrical pulses (action potentials) that travel, via the auditory nerve (Chapter 109), to the cochlear nucleus, the first station of the auditory central nervous system (Chapters 110 and 111).

At several stages of the auditory periphery signal transmission proceeds not only inward but also outward. In Fig. 1, outward-pointing thin arrows indicate that the ear broadcasts sound into the environment. The sound emanates from the cochlea, where the outer hair cells behave as motors that enhance basilar membrane motion.

2 BASILAR MEMBRANE MOTION

2.1 Traveling Wave and Spatial Frequency Analysis

When the middle ear ossicles vibrate, they generate pressure waves that propagate very rapidly (in a few microseconds) throughout the cochlear fluids but also cause a secondary and much slower (transverse) displacement wave that travels on the basilar membrane, from the base of the cochlea toward its apex, in a few milliseconds[3] (Fig. 2*a*). Although the usual physiological route for airborne sound is through vibration of the oval window by the stapes, identical base-to-apex propagation patterns of the displacement wave are generated when sound stimulates the cochlea via an artificial apical window or via bone conduction. The speed of travel of the displacement wave is not uniform as a function of distance: At the basal end propagation is fast and, therefore, for sinusoidal stimulation, the wavelength is long; as the wave proceeds toward the apex, it slows and the wavelength shortens (Fig. 2*a*). The finite speed of travel implies an accumulating delay as a function of distance from the stapes; for sinusoidal stimulation this translates into accumulating phase lag (Fig. 2*b*). Regardless of stimulus frequency, the shape of the vibration envelope is similar in that the basal portion (corresponding to the region of long wavelengths) has a shallow spatial gradient, there is a maximum at a particular longitudinal location, and the maximum is followed on the apical side by a relatively steep slope (Fig. 2*a*). The displacement envelopes are confined to progressively more basal locations as fre-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

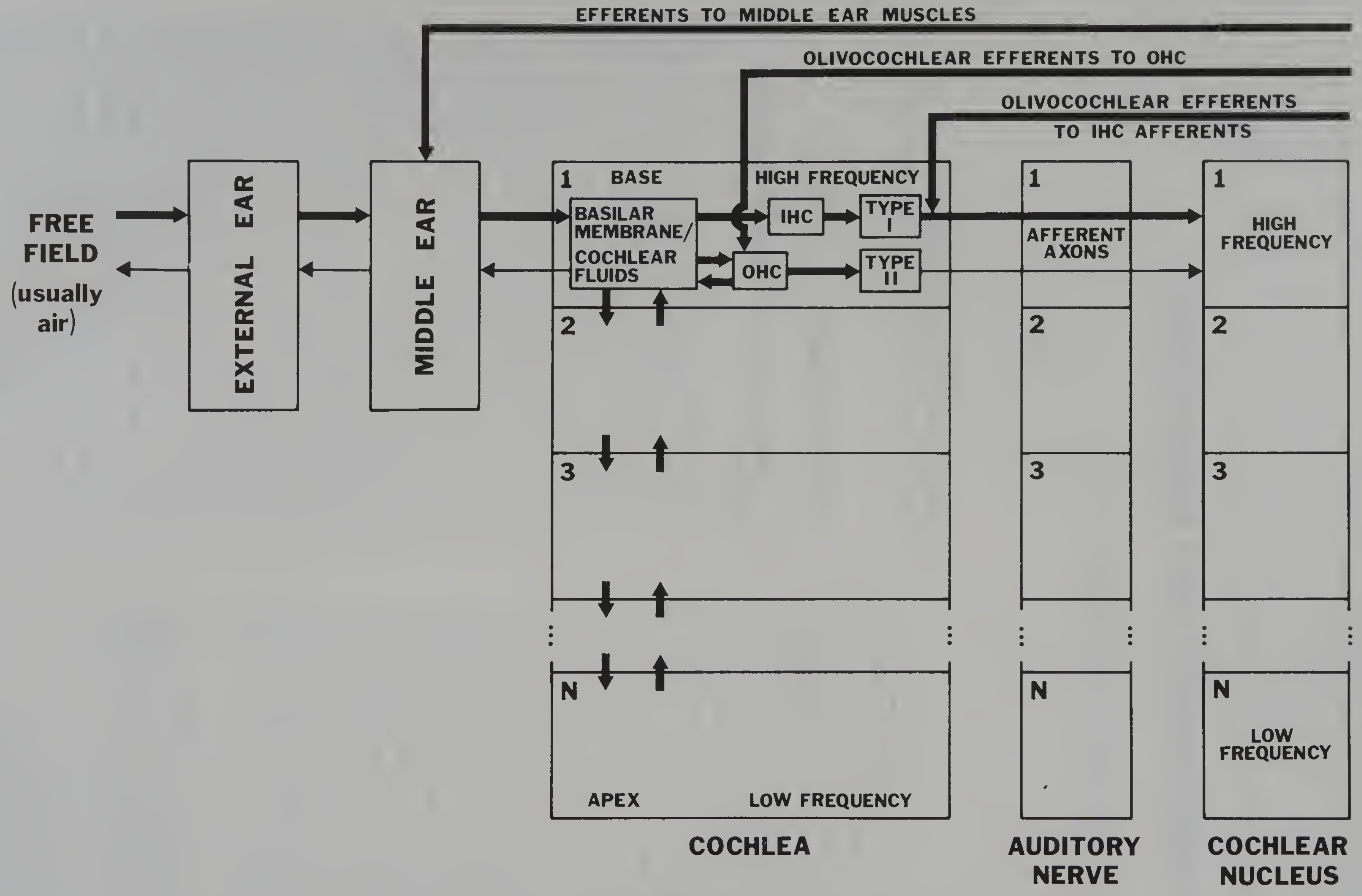

EFFERENTS TO MIDDLE EAR MUSCLES
OLIVOCOCHLEAR EFFERENTS TO OHC
OLIVOCOCHLEAR EFFERENTS TO IHC AFFERENTS
FREE FIELD (usually air)
EXTERNAL EAR
MIDDLE EAR
1
BASE
HIGH FREQUENCY
BASILAR MEMBRANE/ COCHLEAR FLUIDS
IHC
OHC
TYPE I
TYPE II
2
3
N
APEX
LOW FREQUENCY
COCHLEA
1
AFFERENT AXONS
2
3
N
AUDITORY NERVE
1
HIGH FREQUENCY
2
3
N
LOW FREQUENCY
COCHLEAR NUCLEUS

Fig. 1 Block diagram of the peripheral auditory system. Blocks are arranged from peripheral (i.e., closer to the environment, left) to central (i.e., closer to the brain, right). Arrows indicate the direction of signal transmission. Inner hair cells and outer hair cells are represented by the blocks labeled IHC and OHC, respectively. The two types of cochlear afferent neurons are indicated by the blocks labeled Type I and Type II. While the external ear and the middle ear are each represented as a single, undivided box, the cochlea, auditory nerve, and cochlear nucleus are shown subdivided into segments, indicating the systematic mapping of frequency into space. For simplicity, the interconnections within and between the cochlea, the auditory nerve, and the cochlear nucleus are only shown in the segments labeled 1. (Modified with permission from Ref. 2.)

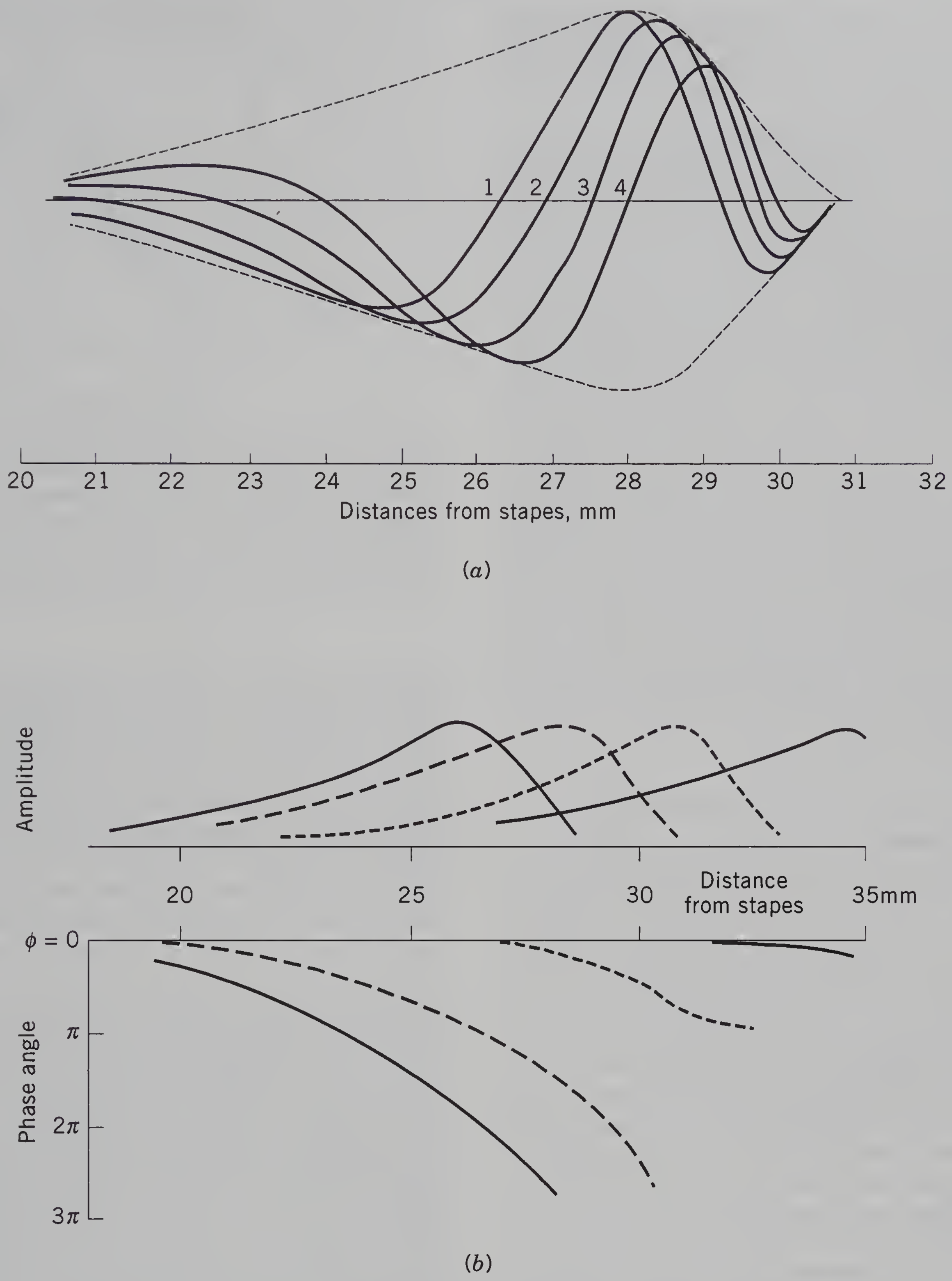


Fig. 2 The basilar membrane traveling wave, as measured by Békésy in human cadavers. (*a*) "Snapshots" at four times (1 = earliest) of the basilar membrane displacement patterns elicited by a 200-Hz tone; the scale of the ordinate is greatly exaggerated relative to that of the abscissa. The dashed lines indicate the peak displacements of the basilar membrane toward scala vestibuli and scala tympani at each cochlear location. (*b*) Peak-to-peak displacements as a function of distance for responses to tones with four different frequencies (upper section), and the corresponding phase patterns (lower section). (Reproduced by permission, with modifications, from Ref. 3.)

quency increases (Fig. 2*b*). Thus, a frequency map is laid out along the cochlea, in which each longitudinal location of the basilar membrane vibrates maximally at its *characteristic frequency* (CF).

The basilar membrane is stiffest at the cochlear base, where it is narrow and thick, and becomes progressively more elastic toward its (wider and thinner) apical end.[3] The variation of basilar membrane elasticity as a function of position (amounting to a ratio of 1 to 100 between the extremes of the human cochlea) is thought to be the fundamental physical characteristic that generates frequency tuning and spatial frequency mapping in the cochlea. Accordingly, mathematical and hardware models of the cochlea that consist of little more than two fluid-filled chambers separated by a membrane with smoothly varying elasticity are able to mimic, at least qualitatively, the main features of basilar membrane motion observed in the ears of dead animals and human cadavers (reviewed in Ref. 4). Such models appropriately exhibit displacement waves that travel from base to apex and whose envelopes are tonotopically organized, with high-frequency responses localized to the stiffer end. However, these simple models are incapable of reproducing realistic basilar membrane responses in healthy cochleae (reviewed in Ref. 5).

2.2 Nonlinearity and Lability of Basilar Membrane Motion

The basilar membrane responds to stimulation by single tones with a transverse AC sinusoidal vibration. At the basal region of normal cochleae, basilar membrane responses to near-CF tones grow with stimulus intensity at rates lower than linear, while response growth is linear at other frequencies[6–8] (reviewed in Ref. 9). Because of the frequency-selective nature of the compressive nonlinearity, basilar membrane responses are more sharply tuned for low-level than for high-level stimuli. Figure 3 shows response magnitude curves, as a function of stimulus frequency and intensity, for a healthy chinchilla basilar membrane site located 3.5 mm from the basal end. To facilitate comparison of tuning and sensitivity as a function of stimulus level, the curves have been normalized to stimulus magnitude; that is, they are presented as gain curves, with the parameter being stimulus level. Responses at frequencies well below CF are linear and remain constant as stimulus level is varied. At low stimulus levels CF responses are sharply tuned and sensitive, but at intense stimulus levels they become insensitive and frequency tuning deteriorates. The normal frequency selectivity of the responses of inner and outer hair cells (Section 3.3) and auditory-nerve fibers (Chapter 109) derives, largely or entirely, from mechanical (i.e., basilar membrane) frequency tuning. This is illustrated in Fig. 4, which compares the frequency tuning of the basilar membrane at a site 3 mm from the basal end of the guinea pig cochlea with that of a cochlear afferent with appropriate CF (about 18 kHz). Although it is not yet certain whether neural thresholds more directly reflect velocity or displacement of the basilar membrane, it is unquestionable that, at least in basal regions of the cochlea, the frequency selectivity displayed by cochlear afferents near CF can be fully accounted for by the frequency tuning of basilar membrane responses. Comparison between neural and mechanical tuning curves for the guinea pig 18-kHz cochlear site suggests that at neural CF threshold the basilar membrane peak displacement is about 0.35 nm, which corresponds to a velocity of 40 μm/s; for the chinchilla 8- to 10-kHz basilar membrane site, the values are 2 nm or 100 μm/s.

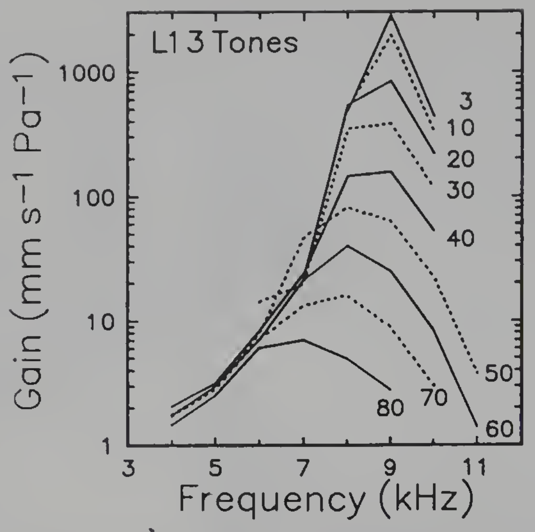


Fig. 3 Frequency response of the chinchilla basilar membrane at a site 3.5 mm from the cochlear base. Gains (i.e., velocity magnitudes divided by pressure magnitudes) are expressed in millimeters per second per Pascal (mm/s/Pa). The parameter is stimulus intensity, expressed in decibels re 20 μPa. (Modified with permission from Ref. 10.)

In the same manner that the sharp tuning of cochlear afferent responses reflects the frequency selectivity of the basilar membrane, mechanical counterparts of CF-specific nonlinearities in auditory-nerve responses to two-tone stimuli also exist at the level of the basilar membrane (reviewed in Ref. 10). Rate suppression in the auditory nerve (see Chapter 109), characterized by the ability of two-tone stimuli to elicit a response rate lower than that evoked by a single tone with frequency near CF, has a clear mechanical analog: The basilar membrane response to a CF tone can be reduced by introduction of a second tone, which by itself elicits smaller responses.[11,12] All the main features of two-tone suppression in the basilar membrane faithfully mirror those of rate suppression in auditory-nerve fibers.

Two-tone distortion products have been unambigu-

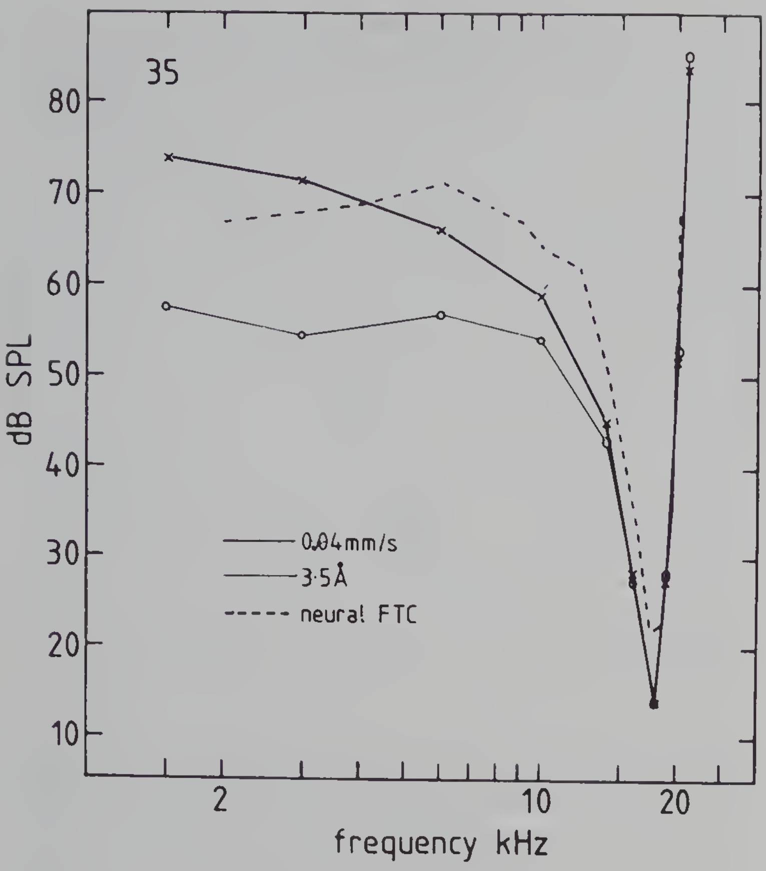


Fig. 4 Comparison of frequency tuning in the basilar membrane and in the auditory nerve. The basilar membrane data, presented as both isovelocity (0.04 mm/s, crosses) and isodisplacement (0.35 nm, open symbols) curves, were recorded from guinea pigs at a site about 3 mm from the basal end of the cochlea. The neural curve (dashed line) is representative of recordings from cochlear afferents with CF similar to that of the basilar membrane recording site. (Reproduced with permission from Ref. 7.)

ously demonstrated in the basilar membrane of several mammalian species.[13–15] Upon stimulation with two primary tones, f_1 and f_2, with frequency near the CF of the basilar membrane site, a variety of combination tones are elicited, with frequencies equal to $2f_1–f_2$, $3f_1–2f_2$, $2f_2–f_1$, and so forth, and magnitudes that vary inversely with separation from CF. At low stimulus levels, distortion product intensity may be no more than 20 dB below primary-tone intensity. If the primary tones are chosen with frequencies such that $2f_1–f_2$ corresponds to CF, it is possible to elicit a distortion product with magnitude far surpassing the response to the primaries. This is strong evidence that the distortion products originate near the sites of the two primary tones and subsequently propagate to the basilar membrane sites with CF corresponding to the distortion product frequency.

Following cochlear damage or death of the animal, basilar membrane responses to tones far from CF remain unchanged while responses to near-CF tones are drastically affected: They become linear and less sensitive, their magnitude being reduced by as much as 40–60 dB[7,8] (reviewed in Ref. 16). Cochlear damage also abolishes two-tone suppression.[12] The fact that the loss of mechanical sensitivity and nonlinearities occurs exclusively at frequencies near CF provides a unifying explanation for the CF specificity of neural two-tone suppression and distortion (Chapter 109) and of the effects of hair cell loss[17] and olivocochlear efferent activation[18] on auditory-nerve responses: All of these effects must reflect alterations of the basilar membrane mechanical response. However, since the basilar membrane is largely acellular and thus probably relatively insensitive to metabolic insults, the lability of its responses suggests that its mechanical properties are influenced by cellular processes.

The sensitivity and frequency tuning of basilar membrane responses to sound can be reversibly reduced by systemic injection of furosemide (Fig. 5), an ototoxic diuretic whose effects on cochlear function result principally from abolishment of the endocochlear potential and a consequent reduction of transduction currents in inner and outer hair cells.[19] The mechanical effect of furosemide strongly implies that the CF-specific sensitivity and nonlinearity of basilar membrane responses depend on hair cell receptor currents or potentials. Further, since the effects of DC currents applied intracellularly to inner hair cells[20] are not CF specific (in contrast with currents applied extracellularly between scala media and scala tympani), the effects of furosemide almost certainly must be mediated by outer hair cells.

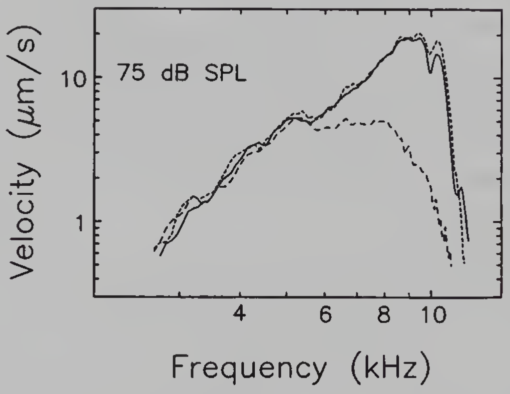


Fig. 5 Effect of a furosemide injection upon the frequency spectrum of basilar membrane responses to clicks. The frequency spectra were obtained by Fourier transformation of velocity responses to 75-dB SPL (sound pressure level re 20 μPa) clicks. The abscissa indicates frequency; the ordinate indicates spectral velocity in units of micrometre per second (μm/s). Three curves are displayed, representing responses immediately preceding (continuous line) and following (long-dash line) a furosemide injection, and when full response recovery had occurred (short dash line). (Reproduced with permission from Ref. 19.)

3 HAIR CELLS

3.1 Apical Membrane: Stereocilia and Transduction

The inner and outer hair cells of the organ of Corti are responsible for the translation of basilar membrane motion into electrical responses in mammalian ears. Hair cells are characterized by apical membrane specializations—rigid microvilli called stereocilia—which underlie their transducer functions (reviewed in Ref. 21). The stereocilia are arranged into a few rows that are graded in height and are aligned into a W or U pattern (see Chapter 107). When the stereocilia are in their normally erect position, a small standing ionic current flows from the scala media into the hair cells through a small percentage of open transduction channels. This current is modulated by deflection of the stereociliar bundle along an axis of maximum sensitivity: Movement toward the tallest row causes an increase in apical membrane conductance, while movement in the opposite direction decreases the conductance.[22,23] Movements of the bundle orthogonal to this axis produce no response, whereas intermediate angles of deflection produce responses whose magnitude depends on the size of the vectoral component along the most sensitive axis. The resultant modulation of the transduction current elicits receptor potentials measurable across the basolateral membrane of the cells; depolarization and hyperpolarization correspond, respectively, to transducer conductance increase and decrease. The changes in membrane conductance are believed to arise from the gating of mechanically sensitive ion channels located near the tips of the stereocilia.[24,25] The gating mechanism may involve direct interactions between interstereociliar microfilaments (so-called tip links[26]) and the transduction channel, which is nonspecifically selective for cations. These tip links, elastic filaments ("springs") that link the top of each stereocilium with a dense membranous plaque on the upper side of the adjacent taller stereocilium, occur in line with the axis of maximal bundle sensitivity. They are believed to provide the tension required to open transduction channels during bundle deflection, one end of the filament being anchored and the other pulling on the channel gate as adjacent stereocilia shear past one another. Since endolymph, the fluid surrounding the stereocilia, has a high K^+ concentration (150 mM), this ion is probably the major charge carrier of the transduction current.

Regardless of sensory organ or animal species, hair cell transduction is nonlinear.[21] Opposite but equal displacements of the hair bundle from the resting position generate unequal conductance changes, resulting in asymmetrical voltage responses. The response to bundle movement toward the tallest stereociliar row, that is, depolarization, is greater (Fig. 6). This asymmetry is highly significant because the depolarizing receptor potentials induce a Ca^{2+} influx, which in turn leads to neurotransmitter release and neural excitation (Chapter 109).

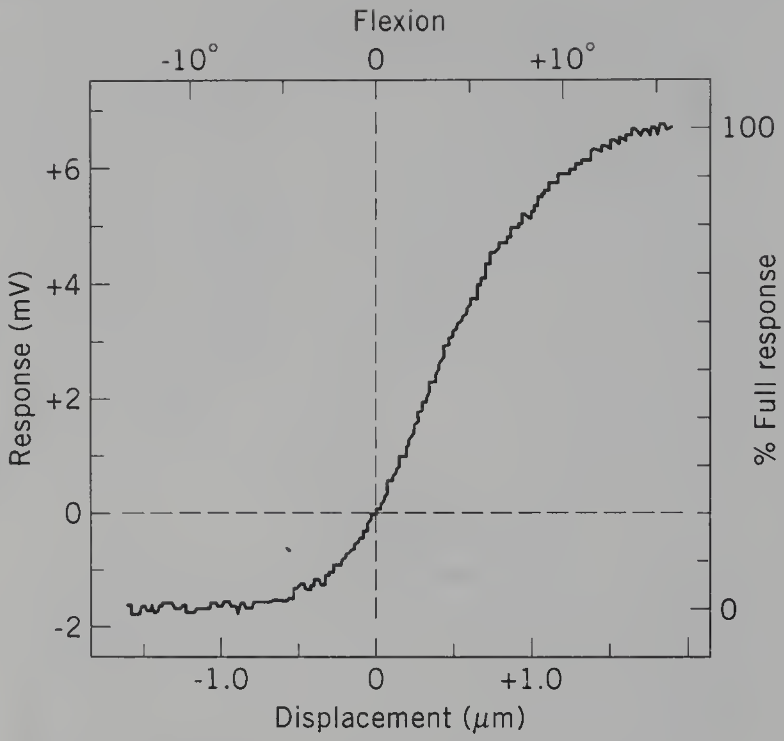


Fig. 6 Sigmoidal stereociliar displacement-response function of hair cell. Due to the offset of the function along the displacement axis, symmetrical sinusoidal displacement of stereocilia produces AC and DC receptor potential components. (Reproduced with permission from Ref. 27.)

3.2 Basolateral Membrane

Hair cell receptor potentials evoked by stereociliar transduction are heavily influenced by the properties of the basolateral membrane, namely, its passive resistance and capacitance and the activity of ligand- and voltage-gated ion channels. Whereas it is possible that the only ion channel in the apical membrane is the one involved in transduction, several types of channels exist in the basolateral membrane. Their conductances have been most completely described in hair cells of nonmammalian species (e.g., those from the frog sacculus and the turtle and chick basilar papillae; reviewed in Ref. 28).

Common among all hair cell types are large outward K^+ currents that activate at membrane potentials more positive than −60 to −50 mV. In vivo, the normal resting potential of the inner hair cell is near −40 mV and the outer hair cell near −70 mV. In inner hair cells isolated from the guinea pig, two types of potassium conductances have been described.[29] The larger of the two has very rapid onset kinetics, approaching a 150-μs onset time constant at large depolarizations. While depolarization induces an inward Ca^{2+} current in inner hair cells, the K^+ currents do not require extracellular Ca^{2+}.

Outer hair cells also display depolarization-induced outward K^+ currents, but their kinetics are much slower than the dominant IHC current: Peak currents occur from 20 to 40 ms after a voltage step.[30,31] There exist at least two K^+ channels.[30] The larger conductance channel (230 pS) is activated at voltages more positive than −35 mV. The smaller (45 pS) is probably active at near-resting membrane potentials (−90 to −50 mV).[32] A non-inactivating inward Ca^{2+} current is evoked by depolarizations greater than −40 mV.[31,33,34] Although both the K^+ channels appear to be Ca-dependent,[30] the bulk of depolarization-induced outward K^+ currents is unaffected by removal of extracellular Ca^{2+}.[33] In addition to voltage-activated ion channels, OHCs possess at least two ligand-activated currents: an ACh-induced slowly activating K^+ current that is dependent on extracellular Ca^{2+}[35] and an ATP-induced inward cationic current, carried most efficiently by Ca^{2+}.[36,37] The ACh-induced K^+ current probably contributes to the cochlear effects of the medial efferent system, whose terminals synapse at the basal membrane of OHCs. The role of the ATP-induced inward cationic current is obscure.

3.3 In Vivo Responses to Sound

The translation of basilar membrane motion into stereociliar deflection underlies sound reception in the mammal. The traditional view of stereocilia excitation is that bending and shearing occur as a result of transverse displacements of the basilar membrane. Since the tips of OHC stereocilia insert into an acellular superstructure, the tectorial membrane, maximum deflection of the basilar membrane toward scala vestibuli should and does coincide with maximal displacement of the stereociliary bundle toward the tallest stereocilia and with maximal OHC depolarization, at least at low stimulus frequencies. However, at high frequencies, differing basilar membrane and outer hair cell responses are measurable and may be attributable to complex micromechanical interactions between stereocilia and the tectorial membrane.[38] Dynamic effects on OHC membrane RC (resistance–capacitance) filter characteristics due to voltage-dependent conductances and capacitance (see below) may also be influential.

The mode of mechanical stimulation of inner hair cells, whose stereocilia may not contact the tectorial membrane, differs from that of the outer hair cell. Inner hair cell stereocilia are thought to be deflected by fluid motion, rather than by shearing against the tectorial membrane.[39] This idea is supported by low-frequency recordings from inner hair cells in both apical and basal cochlear regions.[40,41] However, recording from auditory-nerve afferents, which maintains cochlear integrity, indicates that basal inner hair cells may actually be maximally stimulated when the basilar membrane moves toward scala tympani.[42]

As a consequence of the asymmetrical transducer function of hair cells, sound-induced sinusoidal stimulation of the stereociliary bundle produces a depolarizing DC response in addition to an AC response. This can be seen in the responses of an inner hair cell recorded in vivo from the basal (high-frequency) region of the cochlea (Fig. 7). With increasing stimulus frequency, AC

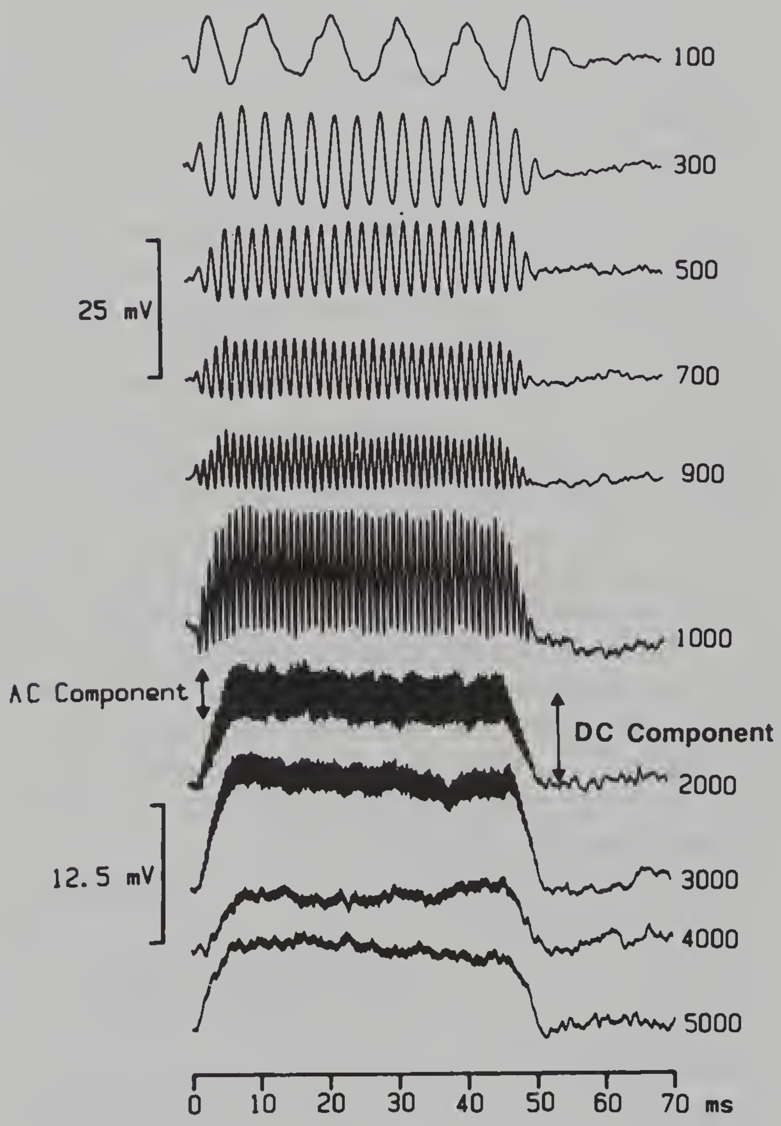


Fig. 7 Inner hair cell receptor potentials in response to tone pips. The hair cell was located at the base of a guinea pig cochlea and had a CF close to that of the tuning curves in Fig. 4. The frequency (in hertz) of the tone-pip stimulus is indicated at the right of each trace. The lower ordinate scale applies to the bottom five traces. Note that the AC component is insignificant in responses to the higher-frequency tone pips. (Reproduced with permission from Ref. 43.)

voltage responses are markedly attenuated by the electrical properties of the basolateral membrane: Its conductances ($1/R$), in combination with its capacitance ($C, 1\mu F/\mathrm{cm}^2$), act as a low-pass frequency filter with a cutoff equal to $\frac{1}{2}\pi$RC (Fig. 7). The frequency extent of phase locking in auditory-nerve fibers (Chapter 109) is probably primarily determined by the roll-off of IHC AC responses,[43] although it may also be influenced by the kinetics that control Ca entry and its intracellular concentration.[44] If it were not for the reduced membrane time constant resulting from the activation of a significant proportion of basolateral conductances at the IHC's resting potential,[29] phase locking would probably roll off at even lower frequencies.

Despite the frequency-dependent roll-off of AC responses, the asymmetrical transducer function of IHCs ensures that high-frequency acoustic information is signaled as DC responses. The polarity of the DC component of IHCs is always positive (excitatory) regardless of frequency of stimulation or location of the cell along the cochlear duct.[45,46] In contrast, the sign of the DC receptor potential of OHCs depends on frequency relative to CF and on stimulus magnitude. The DC response of apical OHCs is hyperpolarizing at frequencies below CF at low stimulus intensities but reverses polarity as intensity increases.[46] The DC responses of basal OHCs to low-frequency stimulation also show a level-dependent sign reversal, but the reversal occurs at much higher stimulus levels than in apical cells.[47] It is puzzling that DC responses of basal OHCs to CF tones can only be measured at very intense sound levels.

In vivo responses of mammalian hair cells show no evidence of adaptation, with receptor potentials remaining uniform in magnitude throughout the duration of acoustic stimulation.[45] This contrasts with transduction in nonmammalian hair cells, where adaptation is ubiquitous.[48–50] However, in vitro, receptor potentials from outer hair cells do exhibit adaptation during static displacements of their stereocilia.[51,52]

At any particular cochlear location, the frequency tuning of inner and outer hair cells is similar[45,46,53] and closely resembles the tuning displayed in the responses of auditory nerve fibers and the basilar membrane. Furthermore, most, probably all, CF-specific response properties of auditory-nerve fibers have clear counterparts in the voltage responses of inner hair cells (Chapter 109).

A variety of experimental manipulations, including electrical stimulation of the olivocochlear efferent system[54] and injection of electrical currents into the cochlea,[20] alters the properties of inner hair cell responses to sound. The effects of these manipulations indicate that outer hair cells are able to influence inner hair cells, probably via an effect on basilar membrane vibration (see Section 2.2).

4 COCHLEA AS A SOUND GENERATOR

4.1 Otoacoustic Emissions

The discovery by Kemp[55] of otoacoustic emissions—that is, sounds produced by the cochlea—was one of the central developments that revolutionized our understanding of inner-ear physiology during the last decade. Kemp showed that when the normal human ear is exposed to clicks or short tone pips, the ear responds, after a fairly long delay (about 10–15 ms), by broadcasting low-frequency (0.5–4 kHz) sounds (Fig. 8). Click-evoked otoacoustic emissions, whose frequency spectra contain several peaks, are unique to each ear and remain remarkably stable over time.

In addition to the click-evoked emissions first measured by Kemp, four other types of otoacoustic emissions can be recognized according to their mode of stimulation (reviewed in Ref. 57). Stimulus-frequency otoacoustic emissions, distortion-product otoacoustic emissions, and electrically evoked otoacoustic emissions are elicited, respectively, by single continuous tones, pairs of tones, and electrical AC currents. Spontaneous otoacoustic emissions[58] are narrow-band sounds that emanate continuously from the ears of a substantial proportion (about 30%) of humans and other primates, and also from the ears of some other animals. While it is doubtful that spontaneous otoacoustic emissions play any role in normal hearing, their existence is highly significant: It demonstrates that, in the absence of external energy inputs (i.e., acoustic stimuli), the cochlea can transduce biological (biochemical) energy into mechanical vibrations. Electrically evoked otoacoustic emissions are produced by injecting AC electrical currents into scala media.[59] These otoacoustic emissions, acoustic analogs of the electrical stimulus, can interact with simultaneously presented external acoustic stimulation. For example, a low-frequency electric current f_{el} concurrent with an acoustic tone f_{ac} can produce otoacoustic emissions at frequencies ($f_{\mathrm{ac}} + f_{\mathrm{el}}$) and ($f_{\mathrm{ac}} - f_{\mathrm{el}}$) as well as at f_{el}. Such interactions constitute evidence not only for the existence in the cochlea of a process of reverse transduction, which converts energy from an electrical to a mechanical form, but also suggest that reverse transduction occurs on a cycle-by-cycle basis. In other words, the motile elements in the cochlea appear to act as AC, rather than DC, motors.

A clue to the cellular origin of cochlear reverse transduction is that subjects affected by sensorineural deafness (usually caused by hair cell loss) in a given frequency range lack otoacoustic emissions (OAEs) over that range,[60] suggesting that hair cells are necessary for the generation of otoacoustic emissions. Other clues come from experimental manipulation of forward trans-

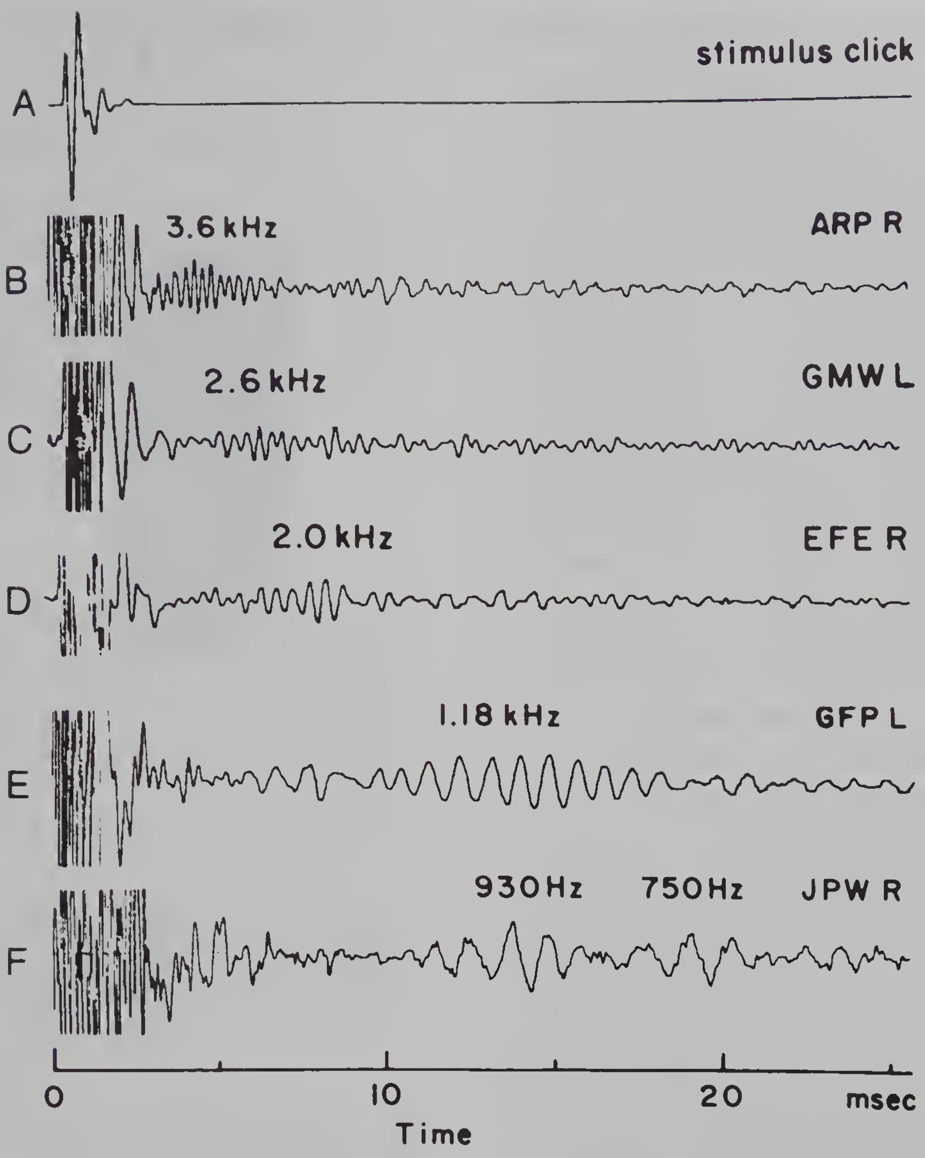


Fig. 8 Click-evoked otoacoustic emissions from human ears. The upper trace indicates the time waveform of the stimulus. The lower five traces show averages of the acoustic waveforms recorded by microphones inserted in the external ear canals of the subjects (identified by name initials and L or R for left and right ear, respectively). Note that each ear produces different emissions and that lower-frequency emissions have longer latencies. (Reproduced with permission from Ref. 56.)

duction in hair cells. Most types of OAEs can be suppressed by tones in a frequency-selective manner[58] and can be altered by changing the endolymphatic potential[61] or by systemic injection of ototoxic drugs.[62] Most significantly, activation of the medial olivocochlear efferent system can also alter the level of distortion-product otoacoustic emissions.[61] Since the medial olivocochlear efferents terminate mostly upon outer hair cells (Chapter 107), it is concluded that the synaptic activity of the efferent terminals can modulate the organ of Corti vibrations, which give rise to otoacoustic emissions. A corollary is that outer hair cells must give rise to otoacoustic emissions and, therefore, must be capable of generating mechanical forces and/or displacements. However, not all otoacoustic emissions are produced by outer hair cells: Otoacoustic emissions can be recorded from the ears of many nonmammalian vertebrates (including frogs, reptiles, and birds) that lack outer hair cells.[63–65]

4.2 Outer Hair Cell Motility

In 1983, Brownell discovered that isolated OHCs are capable of electrically evoked length changes. This form of electrical-to-mechanical transduction may underlie the exquisite sensitivity and frequency selectivity of basilar membrane responses to sound. Brownell et al.[66] determined that transcellular or intracellular current stimulation causes a change in the shape of the OHC; depolarizing currents shorten the cell and hyperpolarizing currents elongate the cell. The mechanical response is not based on a musclelike mechanism, since it does not directly require metabolic substrates (e.g., ATP) or calcium.[30,31,67,68] Nor are cytoskeletal elements involved.[68]

Although the underlying mechanism of OHC electrically evoked motility is unknown, it is clear that its effective stimulus is the transmembrane potential, rather than current.[31,69,70,71] The voltage-dependent nature of OHC motility and the existence of a nonlinear charge movement (up to 2.5 pC per OHC or equivalently a voltage-dependent capacitance of 17 pF) indicate that the mechanical response is dependent on discrete sensor-motor elements within the plasma membrane.[70] The number of voltage sensors within the plasma membrane is estimated to be about 4000–7500/μm^2, similar to the number of intramembranous 10-nm particles observed with electron microscopy.[72–74] Indeed, Dallos et al.[75] have shown that the total length change is the sum of many independent elements along the basolateral surface of the cell. The voltage sensors likewise have been shown to be distributed within the central extent of the cell's length.[76] Furthermore, both sensor and motor reside solely within the plasma membrane.[77,78]

The voltage-to-movement function is sigmoidal[79,80] and begins to saturate in the depolarizing direction at voltages well above physiologically meaningful values.[31] However, saturation in the hyperpolarizing direction occurs near normal in vivo resting potentials (Fig. 9). Consequently, whereas mechanical responses as large as 30 nm/mV have been observed,[31] responses occurring at physiological potentials are much smaller.[79] The significance of this asymmetry is similar to that of the hair cell transducer asymmetry—sinusoidal voltage stimulation at the resting potential produces AC and DC mechanical response components. The DC component is in the contraction-depolarizing direction. Unlike the DC component of the receptor potential, which is unaffected by the RC time constant of the basolateral membrane of the cell, the mechanical DC component is immensely, though indirectly, vulnerable to the effects of the cell's low-pass characteristics.

The physiological consequence of the voltage-dependent nature of OHC motility is substantial. Since receptor potentials drive the mechanical response in vivo, the magnitude and phase of the mechanical response must be governed by the nonlinear RC characteristic of the cell. In vitro, the measured frequency response of motility is limited by the speed of the voltage delivery system.[71]

Fairly flat responses have been measured out to about 20 kHz.[81] Even this value probably underestimates the true mechanical capabilities of the OHC. Nevertheless, estimates of OHC motility derived from measures of OHC receptor potentials in the high-frequency region of the guinea pig cochlea[47] indicate that the AC component would be substantially smaller than basilar membrane motion.[79] The corresponding DC mechanical component would be insignificant.[71] Clearly, it is difficult to reconcile these observations with current concepts of the role that OHCs play in the intact cochlea.

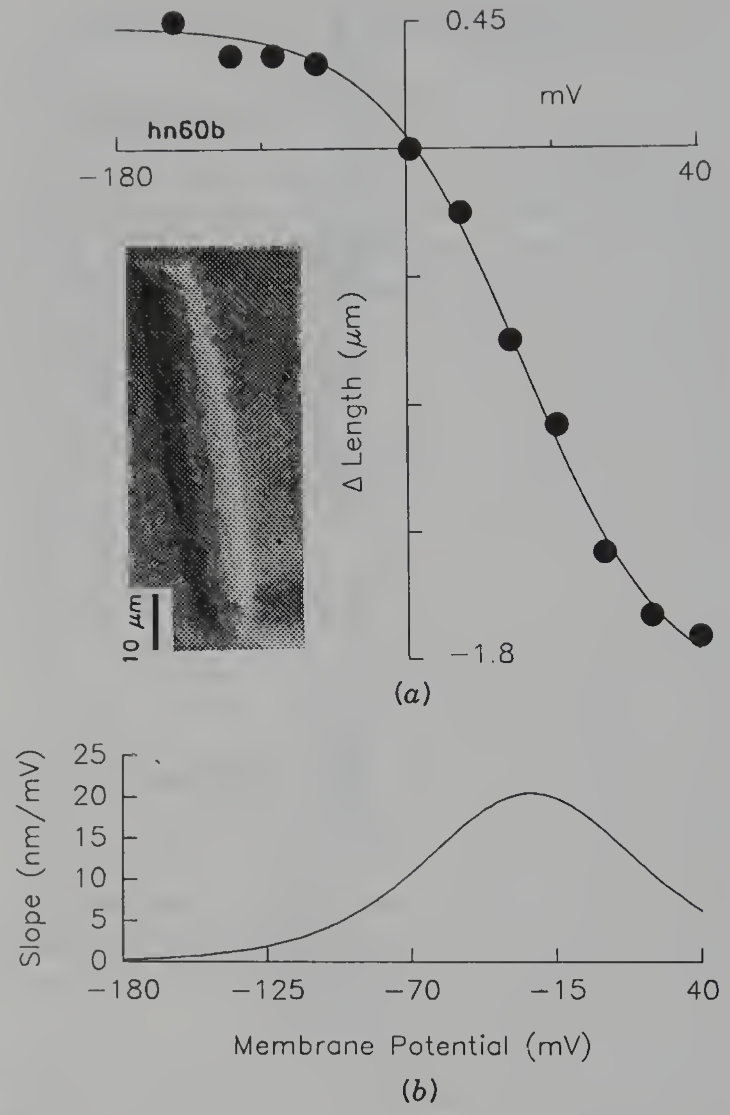


Fig. 9 Mechanical response of mammalian outer hair cell under voltage clamp. (*a*) The OHC changes its length when the cell is held at different membrane potentials. (*b*) The slope of the sigmoidal input–output function defines the cell's responsiveness to membrane potential change. (Reproduced with permission from Ref. 71.)

Nonetheless, the voltage-dependent mechanical response may be dynamically modulated by a variety of physiologically important factors that can modify the mechanical gain of the cell. One of the most effective means of influencing the motility function is by exerting tension on the plasma membrane.[82] Reducing membrane tension increases the gain at the normal resting potential.[83] Since intracellular turgor pressure effectively alters membrane tension, the control of cell turgor may be physiologically significant. Ligand gated ionic channels that can alter the resting membrane potential may also influence the mechanical gain of the cell.[35–37] Finally, it cannot be ruled out that the characteristics of voltage-dependent OHC motility are substantially different in the intact cochlea so as to account for its presumed amplifier activity.

Acknowledgments

M. A. R. was supported by NIH Grants DC-00110 and DC-00419. J. S.-S. was supported by NIH Grants DC-00273 and DC-02003. Many thanks to Nola Rich, John Guinan, and Thomas Weiss for their comments on previous versions of the chapter and to Alberto Recio and Andrei Temchin for help in the preparation of figures.

REFERENCES

1. E. R. Lewis, E. L. Leverenz, and W. S. Bialek, *The Vertebrate Inner Ear*, CRC Press, Boca Raton, FL, 1985.
2. M. A. Ruggero and M. N. Semple, "Acoustics, Physiological," in G. L. Trigg, (Ed.) *Encyclopedia of Applied Physics*, VCH Publishers, Weinheim, Germany, 1991, pp. 213–259.
3. G. von Békésy, *Experiments in Hearing*, McGraw-Hill, New York, 1960.
4. C. D. Geisler, "Mathematical Models of the Mechanics of the Inner Ear," in W. D. Keidel and W. D. Neff (Eds.), *Handbook of Sensory Physiology. Vol. V: Auditory System. Part 3: Clinical and Special Topics*, Springer Verlag, Berlin, 1976, pp. 391–415.
5. E. De Boer, "Auditory Physics. Physical Principles in Hearing Theory. III," *Phys. Rep.*, Vol. 203, 1991, pp. 125–231.
6. W. S. Rhode, "Observations of the Vibration of the Basilar Membrane in Squirrel Monkeys Using the Mössbauer Technique," *J. Acoust. Soc. Am.*, Vol. 49, 1971, pp. 1218–1231.
7. P. M. Sellick, R. Patuzzi, and B. M. Johnstone, "Measurement of Basilar Membrane Motion in the Guinea Pig Using the Mössbauer Technique," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 131–141.
8. L. Robles, M. A. Ruggero, and N. C. Rich, "Basilar Membrane Mechanics at the Base of the Chinchilla Cochlea. I. Input-Output Functions, Tuning Curves, and Response Phases," *J. Acoust. Soc. Am.*, Vol. 80, 1986, pp. 1364–1374.
9. M. A. Ruggero, "Responses to Sound of the Basilar Membrane of the Mammalian Cochlea," *Curr. Opin. Neurobiol.*, Vol. 2, 1992, pp. 449–456.
10. M. A. Ruggero, L. Robles, N. C. Rich, and A. Recio, "Basilar Membrane Responses to Two-Tone and Broadband Stimuli," *Phil. Trans. Roy. Soc. Lond. B*, Vol. 336, 1992, pp. 307–315.
11. R. Patuzzi, P. M. Sellick, and B. M. Johnstone, "The Modulation of the Sensitivity of the Mammalian Cochlea by Low Frequency Tones. III. Basilar Membrane Motion," *Hear. Res.*, Vol. 13, 1984, pp. 19–27.
12. M. A. Ruggero, L. Robles, and N. C. Rich, "Two-Tone Suppression in the Basilar Membrane of the Cochlea: Mechanical Basis of Auditory-Nerve Rate Suppression," *J. Neurophysiol.*, Vol. 68, 1992, pp. 1087–1099.
13. A. L. Nuttall, D. F. Dolan, and G. Avinash, "Measurements of Basilar Membrane Tuning and Distortion with Laser Velocimetry," in P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Eds.), *The Mechanics and Biophysics of Hearing*, Springer Verlag, Berlin, 1990, pp. 288–295.
14. L. Robles, M. A. Ruggero, and N. C. Rich, "Two-Tone Distortion Products in the Basilar Membrane of the Chinchilla Cochlea," in P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Eds.), *The Mechanics and Biophysics of Hearing*, Springer Verlag, Berlin, 1990, pp. 304–311.
15. L. Robles, M. A. Ruggero, and N. C. Rich, "Two-Tone Distortion in the Basilar Membrane of the Cochlea," *Nature*, Vol. 349, 1991, pp. 413–414.
16. M. A. Ruggero, Robles, N. C., L. Robles, and A. Recio, "The Effects of Acoustic Trauma, Other Cochlear Injury and Death on Basilar-Membrane Responses to Sound," in A. Axelsson, H. Borchgrevink, P. A. Hellström, D. Henderson, R. P. Hamernik, and R. Salvi (Eds.), *Scientific Basis of Noise-Induced Hearing Loss*, Thieme Medical Publishers, New York, 1996, pp. 23–35.
17. P. Dallos and D. Harris, "Properties of Auditory Nerve Responses in Absence of Outer Hair Cells," *J. Neurophysiol.*, Vol. 41, 1978, pp. 365–382.
18. M. L. Wiederhold, "Physiology of the Olivocochlear System," in R. A. Altschuler, D. W. Hoffman, and R. P. Bobbin (Eds.), *Neurobiology of Hearing: the Cochlea*, Raven Press, New York, 1986, pp. 349–370.
19. M. A. Ruggero and N. C. Rich, "Furosemide Alters Organ of Corti Mechanics: Evidence for Feedback of Outer Hair Cells upon the Basilar Membrane," *J. Neurosci.*, Vol. 11, 1991, pp. 1057–1067.
20. A. L. Nuttall, "Influence of Direct Current on DC Receptor Potentials from Cochlear Inner Hair Cells in the Guinea Pig," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 165–175.
21. J. Howard, W. M. Roberts, and A. J. Hudspeth, "Mechanoelectrical Transduction by Hair Cells," *Ann. Rev. Biophys. Biophys. Chem.*, Vol. 17, 1988, pp. 99–124.
22. D. P. Corey and A. J. Hudspeth, "Ionic Basis of the Receptor Potential in a Vertebrate Hair Cell," *Nature*, Vol. 281, 1979, pp. 675–677.
23. I. J. Russell, G. P. Richardson, and A. R. Cody, "Mechanosensitivity of Mammalian Auditory Hair Cells In Vitro," *Nature*, Vol. 321, 1986, pp. 517–519.
24. A. J. Hudspeth, "How the Ear's Works Work," *Nature*, Vol. 341, 1989, pp. 397–404.
25. J. O. Pickles and D. P. Corey, "Mechanoelectrical Transduction by Hair Cells," *T.I.N.S.*, Vol. 15, 1992, pp. 254–259.
26. J. O. Pickles, S. D. Comis, and M. P. Osborne, "Cross-

Links between Stereocilia in the Guinea Pig Organ of Corti and their Possible Relation to Sensory Transduction," *Hear. Res.*, Vol. 15, 1984, pp. 103–112.

27. A. J. Hudspeth and D. P. Corey, "Sensitivity, Polarity and Conductance Change in the Response of Vertebrate Hair Cells to Controlled Mechanical Stimuli," *Proc. Natl. Acad. Sci. USA*, Vol. 74, 1977, pp. 2407–2411.
28. P. A. Fuchs, "Ionic Currents in Cochlear Hair Cells," *Prog. Neurobiol.*, Vol. 39, 1992, pp. 493–505.
29. C. J. Kros and A. C. Crawford, "Potassium Currents in Inner Hair Cells Isolated from the Guinea Pig," *J. Physiol. (Lond.)*, Vol. 421, 1990, pp. 263–291.
30. J. F. Ashmore and R. W. Meech, "Ionic Basis of the Resting Potential in Outer Hair Cells Isolated from the Guinea Pig Cochlea," *Nature*, Vol. 322, 1986, pp. 368–371.
31. J. Santos-Sacchi and J. P. Dilger, "Whole Cell Currents and Mechanical Responses of Isolated Outer Hair Cells," *Hear. Res.*, Vol. 35, 1988, pp. 143–150.
32. G. D. Housley and J. F. Ashmore, "Ionic Currents of Outer Hair Cells Isolated from the Guinea-Pig Cochlea," *J. Physiol. (Lond.)*, Vol. 448, 1992, pp. 73–98.
33. J. Santos-Sacchi, "Calcium Currents, Potassium Currents and the Resting Potential in Isolated Outer Hair Cells," Association for Research in Otolaryngology, Midwinter Meeting Abstracts, St. Petersburg, FL, February, Vol. 12, 1989, pp. 81–82.
34. T. Nakagawa, S. Kakehata, N. Akaike, S. Komune, T. Takasaka, and T. Uemura, "Calcium Channel in Isolated Outer Hair Cells of Guinea Pig Cochlea," *Neurosci. Lett.*, Vol. 125, 1991, pp. 81–84.
35. G. D. Housley and J. F. Ashmore, "Direct Measurement of the Action of ACh on Isolated OHCs of the Guinea Pig Cochlea," *Proc. Royal Soc. London*, Vol. 244, 1991, pp. 161–167.
36. T. Nakagawa, N. Akaike, T. Kimitsuki, S. Komune, and T. Arima, "ATP-Induced Current in Isolated Outer Hair Cells of Guinea Pig Cochlea," *J. Neurophysiol.*, Vol. 63, 1990, pp. 1068–1074.
37. J. F. Ashmore and H. Ohmori, "Control of Intracellular Calcium by ATP in Isolated OHCs of the Guinea Pig Cochlea," *J. Physiol. (Lond.)*, Vol. 428, 1990, pp. 109–131.
38. J. J. Zwislocki, "Analysis of Cochlear Mechanics," *Hear. Res.*, Vol. 22, 1986, pp. 155–169.
39. P. Dallos, M. C. Billone, J. D. Durrant, C.-Y. Wang, and S. Raynor, "Cochlear Inner and Outer Hair Cells: Functional Differences," *Science*, Vol. 177, 1972, pp. 356–358.
40. P. Dallos and J. Santos-Sacchi, "AC Receptor Potentials from Hair Cells in the Low Frequency Region of the Guinea Pig Cochlea," in W. R. Webster and L. M. Aitkin, (Eds.), *Mechanisms of Hearing*, Monash University Press, Clayton, Australia, 1983, pp. 11–16.
41. A. L. Nuttall, M. C. Brown, R. I. Masta, and M. Lawrence, "Inner Hair Cell Responses to the Velocity of Basilar Membrane Motion in the Guinea Pig," *Brain Res.*, Vol. 211, 1981, pp. 171–174.
42. M. A. Ruggero, "What Is the Mechanical Stimulus for the Inner Hair Cell? Clues from Auditory-Nerve and Basilar-Membrane Responses to Low-Frequency Sounds," in Å. Flock, D. Ottoson and M. Ulfendahl (Eds.), *Active Hearing*, Elsevier (Pergamon), Oxford, UK, 1995, pp. 321–336.
43. A. R. Palmer and I. J. Russell, "Phase-Locking in the Cochlear Nerve of the Guinea-Pig and Its Relation to the Receptor Potential of Inner Hair Cells," *Hear. Res.*, Vol. 24, 1986, pp. 1–16.
44. R. C. Kidd and T. F. Weiss, "Mechanisms That Degrade Timing Information in the Cochlea," *Hear. Res.*, Vol. 49, 1990, pp. 181–208.
45. I. J. Russell and P. M. Sellick, "Intracellular Studies of Hair Cells in the Mammalian Cochlea," *J. Physiol. (Lond.)*, Vol. 284, 1978, pp. 261–290.
46. P. Dallos, J. Santos-Sacchi, and Å. Flock, "Intracellular Recordings from Outer Hair Cells," *Science*, Vol. 218, 1982, pp. 582–584.
47. I. J. Russell, A. R. Cody, and G. P. Richardson, "The Responses of Inner and Outer Hair Cells in the Basal Turn of the Guinea-Pig Cochlea and in the Mouse Cochlea Grown in vitro," *Hear. Res.*, Vol. 22, 1986, pp. 199–216.
48. R. A. Eatock, D. P. Corey, and A. J. Hudspeth, "Adaptation of Mechanoelectrical Transduction in Hair Cells of the Bullfrog's Sacculus," *J. Neurosci.*, Vol. 7, 1987, pp. 2821–2836.
49. A. C. Crawford, M. G. Evans, and R. Fettiplace, "Activation and Adaptation of Transducer Currents in Turtle Hair Cells," *J. Physiol. (Lond.)*, Vol. 419, 1989, pp. 405–434.
50. N. Hacohen, J. A. Assad, W. J. Smith, and D. P. Corey, "Regulation of Tension on Hair-Cell Transduction Channels: Displacement and Calcium Dependence," *J. Neurosci.*, Vol. 9, 1989, pp. 3988–3997.
51. I. J. Russell, G. P. Richardson, and M. Kössl, "The Responses of Cochlear Hair Cells to Tonic Displacements of the Sensory Hair Bundle," *Hear. Res.*, Vol. 43, 1989, pp. 55–70.
52. J. F. Ashmore, "The Cellular Machinery of the Cochlea," *Exp. Physiol.*, Vol. 79, 1994, pp. 113–134.
53. A. R. Cody and I. J. Russell, "The Responses of Hair Cells in the Basal Turn of the Guinea Pig Cochlea to Tones," *J. Physiol. (Lond.)*, Vol. 383, 1987, pp. 551–569.
54. M. C. Brown and A. L. Nuttall, "Efferent Control of Cochlear Inner Hair Cell Responses in the Guinea Pig," *J. Physiol. (Lond.)*, Vol. 354, 1984, pp. 625–646.
55. D. T. Kemp, "Stimulated Acoustic Emissions from within the Auditory System," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 1386–1391.
56. J. P. Wilson, "Evidence for a Cochlear Origin for Acoustic Re-emissions, Threshold Fine Structure and Tonal Tinnitus," *Hear. Res.*, Vol. 2, 1980, pp. 233–252.
57. R. Probst, B. L. Lonsbury-Martin, and G. K. Martin, "A Review of Otoacoustic Emissions," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 2027–2067.
58. D. T. Kemp, "Evidence of Mechanical Nonlinearity and

Frequency-Selective Wave Amplification in the Cochlea," *Arch. Otorhinolaryngol.*, Vol. 224, 1979, pp. 37–45.

59. A. E. Hubbard and D. C. Mountain, "Alternating Current Delivered into the Scala Media Alters Sound Pressure at the Eardrum," *Science*, Vol. 222, 1983, pp. 510–512.
60. P. M. Zurek, W. W. Clark, and D. O. Kim, "The Behavior of Distortion Products in the Ear Canals of Chinchillas with Normal or Damaged Ears," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 774–780.
61. D. C. Mountain, "Changes in Endolymphatic Potential and Crossed Olivocochlear Bundle Stimulation Alter Cochlear Mechanics," *Science*, Vol. 210, 1980, pp. 71–72.
62. S. D. Anderson and D. T. Kemp, "The Evoked Cochlear Mechanical Response in Laboratory Primates," *Arch. Otorhinolaryngol.*, Vol. 224, 1979, pp. 47–54.
63. G. Manley, M. Schulze, and H. Oeckinghaus, "Otoacoustic Emissions in a Song Bird," *Hear. Res.*, Vol. 26, 1987, pp. 257–266.
64. A. R. Palmer and J. P. Wilson, "Spontaneous and Evoked Otoacoustic Emissions in the Frog Rana Esculenta," *J. Physiol. (Lond.)*, Vol. 324, 1981, p. 64P.
65. C. Köppl and G. A. Manley, "Spontaneous Otoacoustic Emissions in the Bobtail Lizard. 1. General Characteristics," *Hear. Res.*, Vol. 71, 1993, pp. 157–169.
66. W. E. Brownell, C. R. Bader, D. Bertrand, and Y. De Ribaupierre, "Evoked Mechanical Response of Isolated Cochlear Outer Hair Cells," *Science*, Vol. 227, 1985, pp. 194–196.
67. B. Kachar, W. E. Brownell, R. Altschuler, and J. Fex, "Electrokinetic Shape Changes of Cochlear Outer Hair Cells," *Nature*, Vol. 322, 1986, pp. 365–368.
68. M. C. Holley and J. F. Ashmore, "On the Mechanism of a High Frequency Force Generator in Outer Hair Cells Isolated from the Guinea Pig Cochlea," *Proc. R. Soc. Lond. B*, Vol. 232, 1988, 413–429.
69. J. F. Ashmore, "Transducer Motor Coupling in Cochlear Outer Hair Cells," in J. P. Wilson and D. J. Kemp (Eds.), *Cochlear Mechanisms: Structure, Function and Models*, Plenum Press, New York, 1989, pp. 107–113.
70. J. Santos-Sacchi, "Reversible Inhibition of Voltage Dependent Outer Hair Cell Motility and Capacitance," *J. Neurosci.*, Vol. 11, 1991, pp. 3096–3110.
71. J. Santos-Sacchi, "On the Frequency Limit and Phase of Outer Hair Cell Motility: Effects of the Membrane Filter," *J. Neurosci.*, Vol. 12, 1992, 1906–1916.
72. R. L. Gulley and T. S. Reese, "Regional Specialization of the Hair Cell Plasmalemma in the Organ of Corti," *Anat. Rec.*, Vol. 189, 1977, pp. 109–124.
73. A. Forge, "Structural Features of the Lateral Walls in Mammalian Cochlear Outer Hair Cells," *Cell Tiss. Res.*, Vol. 265, 1991, pp. 473–483.
74. M. C. Holley, F. Kalinec, and B. Kachar, "Structure of the Cortical Cytoskeleton in Mammalian Outer Hair Cells," *J. Cell Sci.*, Vol. 102, 1992, pp. 569–580.
75. P. Dallos, B. N. Evans, and R. Hallworth, "On the Nature of the Motor Element in Cochlear Outer Hair Cells," *Nature*, Vol. 350, 1991, pp. 155–157.
76. G.-J. Huang and J. Santos-Sacchi, "Mapping the Distribution of the Outer Hair Cell Motility Voltage Sensor by Electrical Amputation," *Biophys. J.*, Vol. 65, 1993, pp. 2228–2236.
77. F. Kalinec, M. C. Holley, K. H. Iwasa, D. J. Lim, and B. Kachar, "A Membrane-Based Force Generation Mechanism in Auditory Sensory Cells," *Proc. Natl. Acad. Sci. USA*, Vol. 89, 1992, pp. 8671–8675.
78. G.-J. Huang and J. Santos-Sacchi, "Motility Voltage Sensor of the Outer Hair Cell Resides within the Lateral Plasma Membrane," *Proc. Natl. Acad. Sci. USA*, Vol. 91, 1994, pp. 12268–12272.
79. J. Santos-Sacchi, "Asymmetry in Voltage Dependent Movements of Isolated Outer Hair Cells from the Organ of Corti," *J. Neurosci.*, Vol. 9, 1989, pp. 2954–2962.
80. B. N. Evans, P. Dallos, and R. Hallworth, "Asymmetries in Motile Responses of Outer Hair Cells in Simulated In Vivo Conditions," in J. P. Wilson and D. J. Kemp (Eds.), *Cochlear Mechanisms: Structure, Function and Models*, Plenum Press, New York, 1989, pp. 205–206.
81. P. Dallos and B. N. Evans, "High-Frequency Motility of Outer Hair Cells and the Cochlear Amplifier," *Science*, Vol. 267, 1995, pp. 2006–2009.
82. K. H. Iwasa, "Effect of Stress on the Membrane Capacitance of the Auditory Outer Hair Cell," *Biophys. J.*, Vol. 65, 1993, pp. 492–498.
83. S. Kakehata and J. Santos-Sacchi, "Membrane Tension Directly Shifts Voltage Dependence of Outer Hair Cell Motility and Associated Gating Charge," *Biophys. J.*, Vol. 68, 1995, pp. 2190–2197.

109

PHYSIOLOGY OF THE AUDITORY NERVE

D. O. KIM AND K. PARHAM

1 INTRODUCTION

The auditory nerve (AN), which includes thousands of afferent fibers, conveys information encoded in the spatio-temporal patterns of action potentials from the cochlea to the brain. On anatomical grounds (see Chapter 107) the afferent AN fibers have been subdivided into two groups: the majority (90–95%), known as type I, innervate the inner hair cells (IHCs) and the remainder, known as type II, innervate the outer hair cells (OHCs). To our knowledge, little is known about the acoustic response properties of type II fibers[1] and it is likely that all of the single AN fiber recordings reported here and in the literature represent type I afferent fibers.[2] The AN also includes fibers of the efferent olivocochlear bundle (OCB) that transmit from the brain to the cochlea. In this chapter we provide a brief general overview of the physiological characteristics of afferent AN and efferent OCB fibers.

2 FREQUENCY TUNING OF AN FIBERS

Every AN fiber exhibits frequency tuning. One measure of tuning is the frequency-threshold tuning curve, which is a plot of the threshold for pure-tone stimuli versus frequency.[3] The frequency corresponding to the lowest threshold or tip of the tuning curve is called the characteristic frequency (CF) of the fiber. The CF of a fiber is related to the place along the spiral cochlear partition innervated by the fiber according to a systematic frequency place organization,[2] which reflects the underlying frequency place transformation resulting from cochlear mechanics (see Chapter 108). The individual signal-encoding characteristics (e.g., discharge rate vs. stimulus level) of a group of AN fibers of similar CF may be quite distinctive, varying with the spontaneous rate (SR), that is, the discharge rate in the absence of an overt stimulus. Knowledge of the CF and SR of a fiber enables prediction of the overall response characteristics of that fiber.

3 SR DISTRIBUTION OF AN FIBERS

The AN fibers exhibit a bimodal distribution of SR.[4,5] Figure 1 illustrates a sample of AN fibers from a study of unanesthetized decerebrate cats.[6] The trough between the two modes is at approximately 15–25 spikes/s. Reflecting the SR distribution, the AN fibers are often divided into low-SR (LSR), medium-SR (MSR), and high-SR (HSR) groups with SRs of 0–0.5, 0.5–18, and 18–higher spikes/s, respectively. Since the border between the LSR and MSR groups is not clearly recognizable in the SR distribution, some studies[5–8] have divided the AN fibers into only two groups. We will refer to the combined LSR and MSR fibers as the LMSR fibers.

4 POPULATION RESPONSE PROFILES OF AN FIBERS

One approach to the analysis of AN encoding of frequency, intensity, and spectral patterns has involved population studies in which the responses of a large number of fibers to a particular stimulus were recorded sequentially. This approach, introduced by Pfeiffer and Kim,[9] has been taken by a number of studies.[5–7,10–13] Figure 2 shows spatial profiles of discharge rates of LMSR and

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

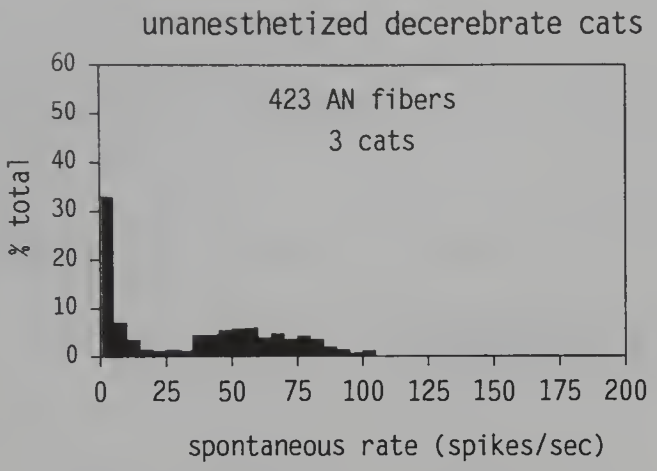


Fig. 1 The distribution of spontaneous rate of auditory nerve (AN) fibers in unanesthetized decerebrate cats. (From Ref. 6.)

HSR AN populations to a 1-kHz pure tone at three different levels in unanesthetized decerebrate cats. At 30 dB SPL re 20 μPa, both LMSR and HSR populations exhibited a peak of response at the characteristic place, i.e., the location at which the CF is equal to the stimulus frequency. As the stimulus level increased to 50 and 70 dB SPL, the HSR profile became increasingly saturated and flattened whereas the LMSR profile retained a more visible peak.

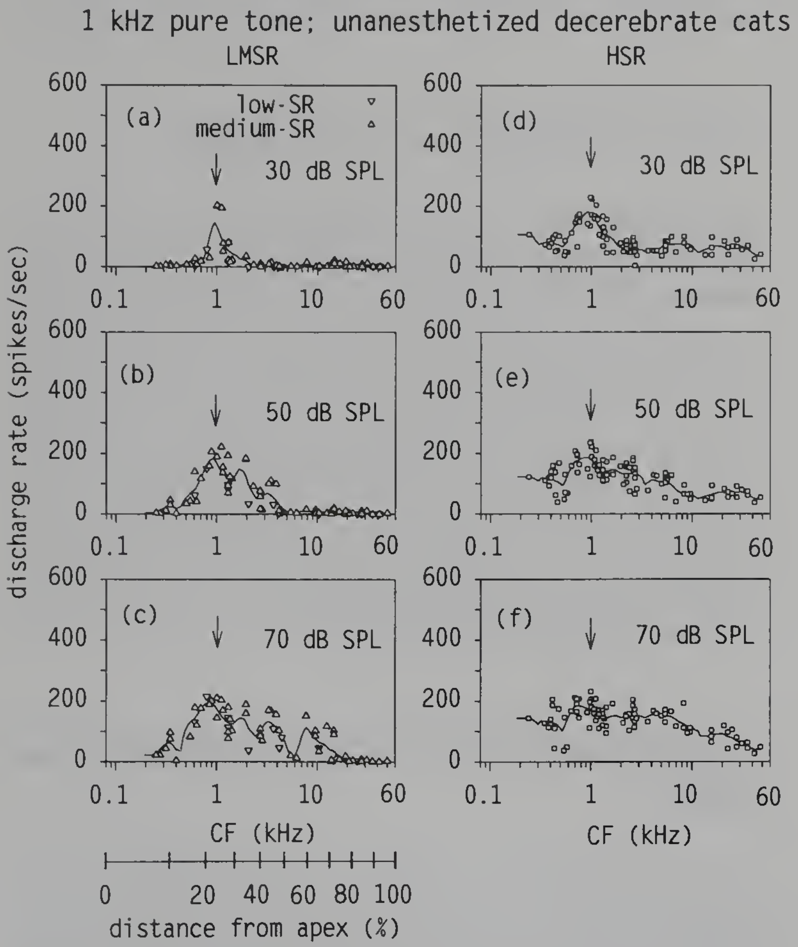


Fig. 2 Spatial response profiles of discharge rates of low and medium spontaneous rate (LMSR, left column), and high-SR (HSR, right column) AN fibers in unanesthetized decerebrate cats to 1 kHz pure tone at 30, 50, and 70 dB SPL. (From Ref. 6.)

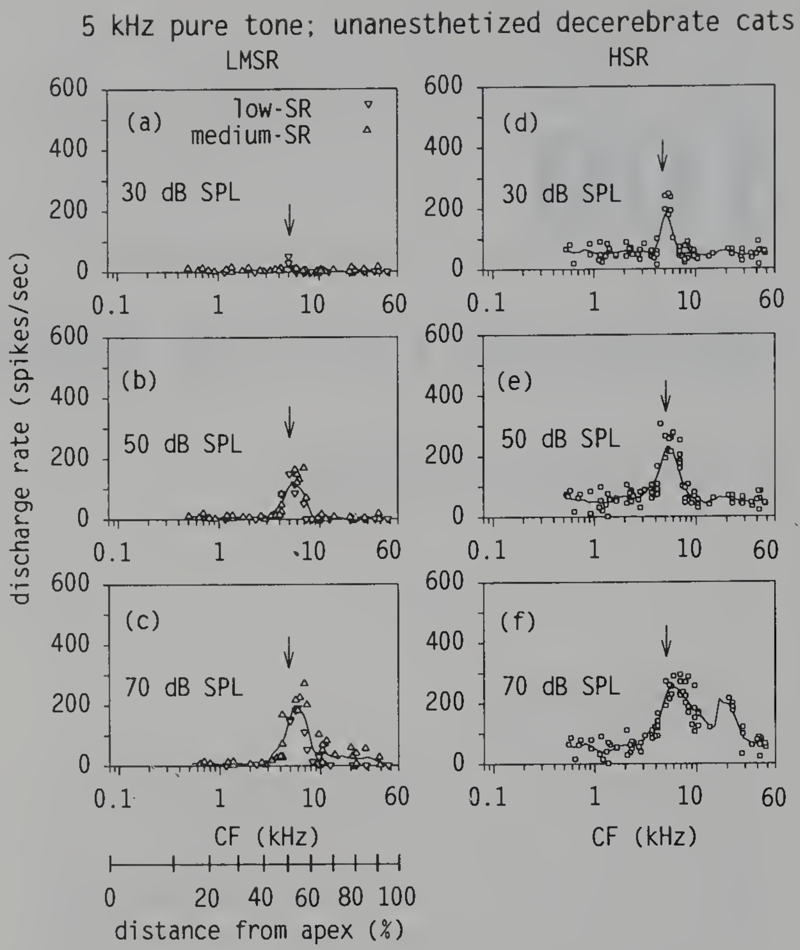


Fig. 3 Spatial response profiles of discharge rates of AN fibers to 5 kHz pure tone. The format of this figure is the same as Fig. 2. (From Ref. 6.)

Figure 3 is analogous to Fig. 2 with the difference being that the stimulus is a 5-kHz pure tone. At 30 dB SPL, the HSR profile exhibited a robust, well-defined peak at the characteristic place whereas the LMSR fibers near the characteristic place were only barely excited. This difference is consistent with the tendency of the thresholds of LMSR fibers to be higher than those of HSR fibers.[4] At 50 dB SPL, both LMSR and HSR populations exhibited a well-defined peak. At 70 dB SPL, the HSR profile exhibited nonlinearity in the form of an abrupt growth of a secondary peak (or bump) at the 15–30 kHz place. A corresponding secondary peak was not clearly visible in the LMSR profile.

5 AN FIBER PLACE CODE OF FREQUENCY

The spatial location (along the cochlear partition) of the peak of each fiber's discharge rate profile, called a rate-place code (or place code), constitutes a potential code for frequency. On the basis of the AN spatial profiles of mean spike count (analogous to Fig. 3) and standard deviation of the spike count in response to a high-frequency pure tone, Kim and Parham[14] analyzed the accuracy of the rate-place code of frequency in terms of the quantity known as *d′* in signal detection theory.[15] Kim and Parham found that, for a high-frequency (5 kHz) pure tone, a rate-place code of frequency in the cat AN provides sufficient information to account for the cat's behavioral frequency discrimination, that is, $\Delta f/f = 0.0088$.[16] For low frequencies (below about 5 kHz), both a place code and a temporal code of frequency are believed to be of potential significance.

6 AN FIBER PHASE LOCKING: A CODE OF FREQUENCY AND SIGNAL TIME/PHASE

A second potential code of frequency besides the place code is a temporal code based on phase locking of AN fiber discharges, that is, the tendency for discharges to occur in a restricted segment (or phase angle) of the cycle of a periodic stimulus. Such phase locking can be discerned in folded period histograms or in interspike interval histograms.[3,17] Phase locking of AN fibers gradually deteriorates with increasing frequency above about 1 kHz and is generally difficult to detect above about 5 kHz.[18] Thus, the phase-locked encoding of frequency is limited to frequencies below about 5 kHz.

Besides possibly encoding stimulus frequency, phase locking of AN fiber discharges fulfills another important function, that is, the encoding of time or phase information about the acoustic signal reaching each of the two ears. Thus, the two auditory nerves arising from the two ears of a listener convey time and/or phase information about the signal in each ear to the brain-stem auditory nuclei. From this information, the brain-stem nuclei can compute interaural time (or phase) difference, a parameter known to be used in discerning the spatial direction of a sound source (see Chapters 110, 111, and 117).

HSR fibers can exhibit phase locking to a low-frequency pure tone at stimulus levels as low as 10–15 dB below that required to produce an increase in mean discharge rate above the spontaneous rate.[17] It has been suggested[6] that phase locking may confer upon HSR fibers a unique advantage over LSR fibers by virtue of the higher spontaneous activity of the former; an HSR fiber can convey information about negative and positive directions of cochlear partition displacement by decreasing as well as increasing their instantaneous rate relative to the SR. In contrast, an LSR fiber can signal only one direction since the instantaneous rate cannot decrease below zero. We postulate that this advantageous phase locking of HSR fibers may be a possible functional reason (from a system design point of view) underlying the dominant projection of HSR AN fibers to certain selected cells of the cochlear nucleus.[19] Such specialized cells may represent a *time-processing channel* as a component in the parallel auditory brain-stem pathways (see Chapters 110 and 111).

7 RATE-LEVEL CHARACTERISTICS AND CODING OF INTENSITY IN AN FIBERS

Figure 4 describes rate-level characteristics of AN fibers. (Note: Figure 4 was not previously published. The methods for obtaining these data are the same as in Kim et al.[6]) Figure 4 represents AN fibers such as those near the peak in Fig. 3 (i.e., with CFs near the stimulus frequency. In Fig. 4, the stimulus frequency was fixed at 5 kHz, and the plots depict all of the fibers sampled over a narrow range of CF near the stimulus frequency. Thus, these plots allow one to compare directly the two SR groups in the same narrow CF range responding to the same stimulus. We can note from Fig. 4 the following differences in the rate-level characteristics of the LMSR and HSR

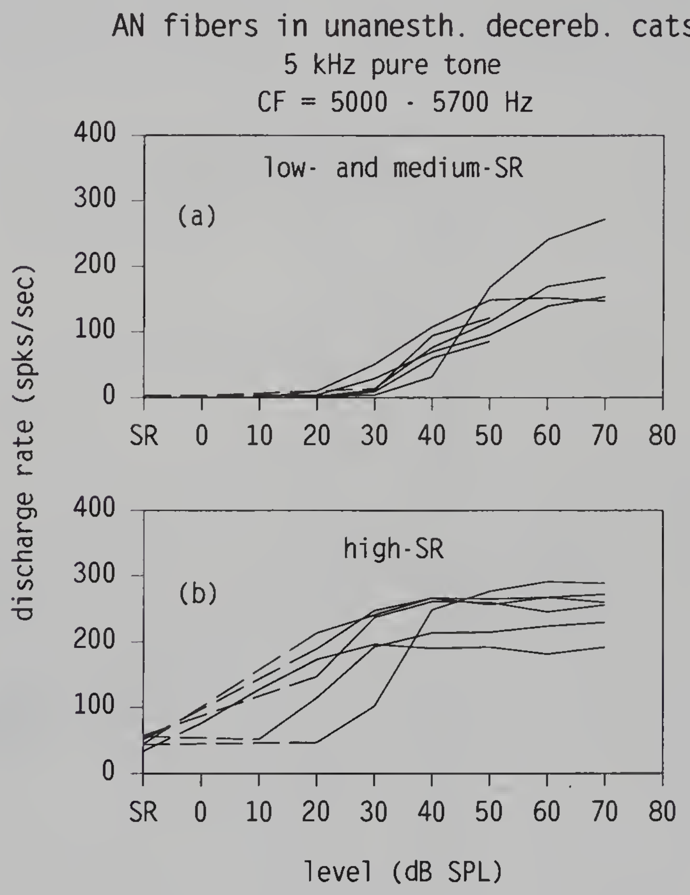


Fig. 4 Rate-level characteristics of low- and medium-SR (top row) and high-SR (bottom row) AN fibers. The plots represent responses to 5 kHz pure tone for all AN fibers sampled in three unanesthetized decerebrate cats, which have characteristic frequencies (CF) in the range of 5.0–5.7 kHz. Each of the lines represents a separate AN fiber. The dashed lines are connections between the discharge rates at the lowest stimulus level tested and the SR. Previously unpublished data.

fibers: (1) Thresholds of LMSR fibers tend to be higher than those of HSR fibers (further discussed in Section 8 below); and (2) the dynamic ranges of LMSR fibers tend to be larger than those of HSR fibers. Other studies of different SR groups of AN fibers are consistent with the above notions regarding the rate-level characteristics[20–22] and threshold vs. CF (see Section 8).

A recent study of Liberman[19] indicated that the AN input to a certain area of the cochlear nucleus (see Chapter 110), that is, the "small cell cap" area, consisted nearly exclusively of LMSR fibers. This unique anatomical feature and the capability of LMSR AN fibers to encode stimulus intensity over a wide dynamic range extending to high stimulus levels suggest that such a pathway may represent an *intensity-processing channel* as another part of the parallel auditory brainstem pathways.

Quantitative evaluation of the ability of AN fibers to encode stimulus intensity (and intensity increments) requires examination of the variability of discharge rate in addition to the mean discharge rate. A number of studies took such an approach.[23,24] Some studies[25–28] represented single unit response in the context of a discrimination task in terms of the d' measure of signal detection theory.[15] From these studies, it is generally believed that LMSR fibers are better able to encode stimulus intensity than HSR fibers. Further studies are needed to provide comprehensive physiological correlates of various psychophysical observations on intensity discrimination (see Chapter 118).

8 THRESHOLDS OF AN FIBERS WITH DIFFERENT SRS

Figure 5 plots thresholds of different SR groups of AN fibers versus CF.[29] (Efferent fiber data in Fig. 5 will be discussed in Section 12.) In Fig. 5, thresholds are defined as the stimulus level that produced a rate increase of 1 spike during a 50-ms tone burst (equivalent to 20 spikes/s rate increase). The HSR AN fibers (×) exhibit the lowest thresholds. The MSR (+) and LSR (☆) fibers exhibit higher thresholds than the HSR fibers across all CFs by about 20 dB or more. A similar finding was reported by Schmiedt.[30] Rather than taking a constant number of spikes/s rate increase as the threshold criterion, other studies have used different criteria. Geisler et al.[31] and Relkin and Pelli[32] used criteria that incorporated both the mean and standard deviaton of spike counts and found that the differences in thresholds between the

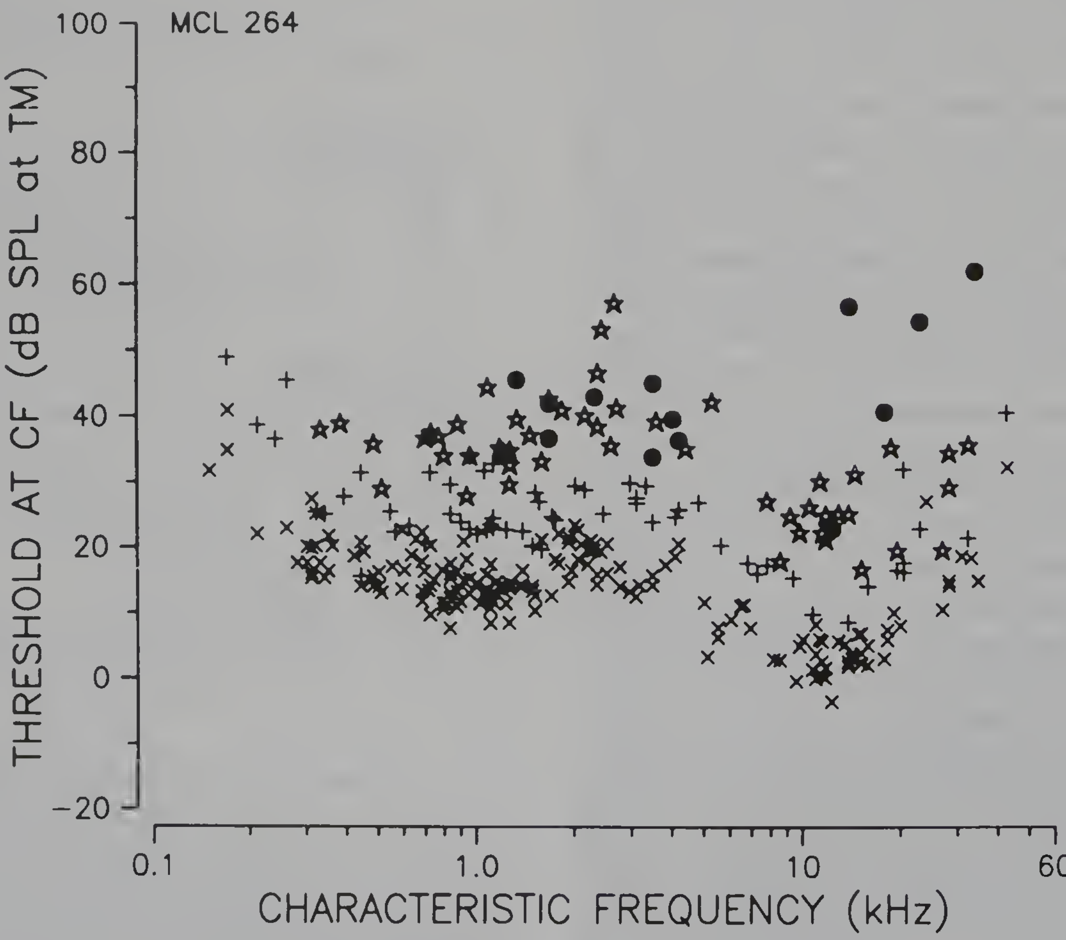


Fig. 5 Thresholds of the afferent LSR, MSR, and HSR AN fibers vs. CF, represented by (☆), (+), and (×), respectively. Thresholds of the efferent olivocochlear bundle (OCB) fibers are represented by (●). (From Ref. 29.)

LMSR and HSR groups were about 5–8 dB, somewhat less than those found by Liberman and by Schmiedt.

9 ADAPTATION OF AN FIBERS

Basic adaptation characteristics of AN fibers are described in the literature.[33–35] An interesting recent finding by Relkin and Doucet[36] is that, in terms of the effects of the interstimulus silent interval in repeated 100-ms tone bursts, the LSR AN fibers require a much longer silent interval to exhibit a fully recovered poststimulus time histogram pattern than the HSR AN fibers, that is, over 1000 ms for the former versus less than 200 ms for the latter. A possible psychophysical correlate of this phenomenon in an intensity discrimination task has been observed by Zeng et al.[37] A more detailed discussion of AN adaptation is covered in Chapter 122.

10 NONLINEAR PHENOMENA IN AN FIBERS

A major new concept arising from recent studies of cochlear biomechanics and the motility of outer hair cells (OHCs) is that the cochlear partition motion is amplified by about 40–60 dB at the CF and the frequency selectivity considerably enhanced by an active nonlinear motile mechanism in the OHCs (see Chapter 108). Several forms of highly nonlinear phenomena present in AN fiber responses are believed to originate from nonlinear cochlear biomechanical mechanisms. The main examples are the generation of propagated distortion products and some aspects of signal suppression.

Studies of distortion products in the response of AN fibers[38–40] and related cochlear modeling studies[41,42] led to the following conclusions. When two tones at similar frequencies, f_1 and f_2, are applied, prominent distortion products, including components at frequencies of $2f_1 - f_2$ and $f_2 - f_1$, are generated in the cochlear region where the f_1 and f_2 vibration patterns overlap. The distortion products are then propagated apically to the distortion frequency place and basally to the stapes, the middle ear, and ear canal. The distortion products are perceptually recognizable as combination tones[43] (see Chapter 114). The appearance of the distortion signals as sounds in the ear canal is called distortion product otoacoustic emissions (see Chapters 108 and 120).

When a second tone is added, the response of an AN fiber to the first tone is often suppressed. This phenomenon is called two-tone suppression[44] observed in two forms: a reduction in mean discharge rate (*two-tone rate suppression*) and in vector strength or synchrony[45] (*two-tone synchrony suppression*). Suppression of a weaker sinusoid by a stronger sinusoid is an inherent property of a compressive nonlinearity.[46] Greenwood[47] showed that a reduction of the vector strength for one sinusoid due to the addition of a second sinusoid can be reproduced even by a model where the only nonlinearity is a half-wave rectifier representing the afferent synapse between an inner hair cell (IHC) and a cochlear neuron. A phenomenological model of the cochlea consisting of a compressive nonlinearity interposed between two linear bandpass filters[48] was successful in reproducing the salient features of two-tone suppression in AN fibers. The presence of the bandpass filter at the output of the nonlinearity is essential to the ability of the model to reproduce suppression of the total output. The biological identity of such an *output bandpass filter* has been difficult to ascertain.

It appears that the overall signal suppression phenomena present in AN fibers are the result of contributions from multiple mechanisms including compressive, frequency-dependent nonlinearity of cochlear partition motion,[49,50] which in turn is expected to depend on OHC's motile mechanisms,[51] nonlinearity in IHC response,[52] and half-wave rectification occurring at the IHC-afferent synapse. Explanation for two-tone rate-suppression phenomenon for a suppressor frequency below the CF of an AN fiber continues to be problematic (see Kim[53] for further discussion). Regardless of the exact mechanisms, it is clear that signal suppression phenomena in AN fibers produce significant effects in AN encoding of complex spectral patterns such as those of vowels as discussed below.

11 REPRESENTATION OF VOWELS IN AN FIBERS

The encoding of spectral patterns of vowels in AN fibers was investigated by Sachs and Young[54] using a population recording method. Their main findings, for the case of [ϵ], were: (1) At about 38–58 dB SPL, peaks in the rate-place profiles of both HSR and LSR fibers in response to a vowel stimulus conveyed formant patterns of the vowel; (2) at about 68–78 dB SPL, the profiles of HSR fibers were heavily saturated and flattened whereas those of LSR fibers maintained peaks conveying the formant patterns; and (3) at 28 dB SPL, the profiles of HSR fibers conveyed the formant patterns and the LSR data were unavailable for the low level. The AN responses to vowels were much affected by signal suppression effects. Sachs and Young[54] reported, however, that the observed effects were not to sharpen the formant representation. Their results with other vowels were consistent with the above.

In a related study, Young and Sachs[55] analyzed AN fiber responses to vowels in terms of a phase-locking measure (i.e., Fourier components of period histograms) and defined a measure called *average localized synchronized rate* (ALSR). This study demonstrated that the ALSR conveyed the formant patterns stably over a wide range of levels. Whether the perceptual recognition of vowels is mediated by the temporal phase-locking representation of AN fibers (such as the ALSR) or the rate-place representation is an open question.

12 PHYSIOLOGICAL CHARACTERISTICS OF EFFERENT OCB FIBERS

Efferent fiber discharges exhibit relatively constant interspike intervals,[56] unlike afferent AN fibers that discharge a Poisson-like random train of spikes.[18] Single efferent fibers are classified according to their responses to monaural tone bursts applied to the ipsilateral or contralateral ear. Approximately two-thirds of the efferent fibers sampled responded best to ipsilateral tones and the remaining one-third to contralateral.[57] A large majority of efferent fibers exhibited binaural facilitation. The addition of noise or tone in the opposite ear, which were not excitatory by themselves, often enhanced the response to the main-ear stimulation.

Thresholds of ipsilateral efferent fibers, together with those of afferent AN fibers in the same animals, were investigated by Liberman.[29] Figure 5 shows these thresholds versus CF where the efferent fibers are denoted with the symbol ●. Efferent fiber thresholds are typically in the range of 30–60 dB SPL and often overlap with those of the LSR afferent fibers. The rate-level characteristics of an ipsilateral efferent fiber are shown in Fig. 6*a*. The solid and dashed lines represent the cases of pure tone stimulation in the ipsilateral ear, with and without a background noise applied to the contralateral ear. Noise in the contralateral ear had a big effect; the threshold was noticeably reduced and the overall discharge rate increased. In the top two curves of Fig. 6*a*, the fiber exhibited a very large dynamic range, 80 dB, without any sign of rate saturation. Figure 6*b* shows the latency-level counterpart of the efferent fiber data presented in Fig. 6*a*. When pure tones were presented alone, the latency was in the 20–50 ms range showing a decrease in latency with an increase in stimulus level. Noise in the contralateral ear dramatically reduced the latency to about 6–10 ms in some units throughout the entire range of stimulus level tested, 10–90 dB SPL.

Activation of the efferent bundle leads to a reduction in the afferent AN response. This was observed in the cochlear compound action potentials[58] and in single units of the AN.[56,59] Such an effect is believed to represent a reduction in the gain of the "cochlear amplifier" mediated by motile activities of the OHCs (see Ref. 60 and Chapter 108). Winslow and Sachs[61] observed that an efferent activation led to an improvement of signal-to-noise ratio in a condition where discharge rates of AN

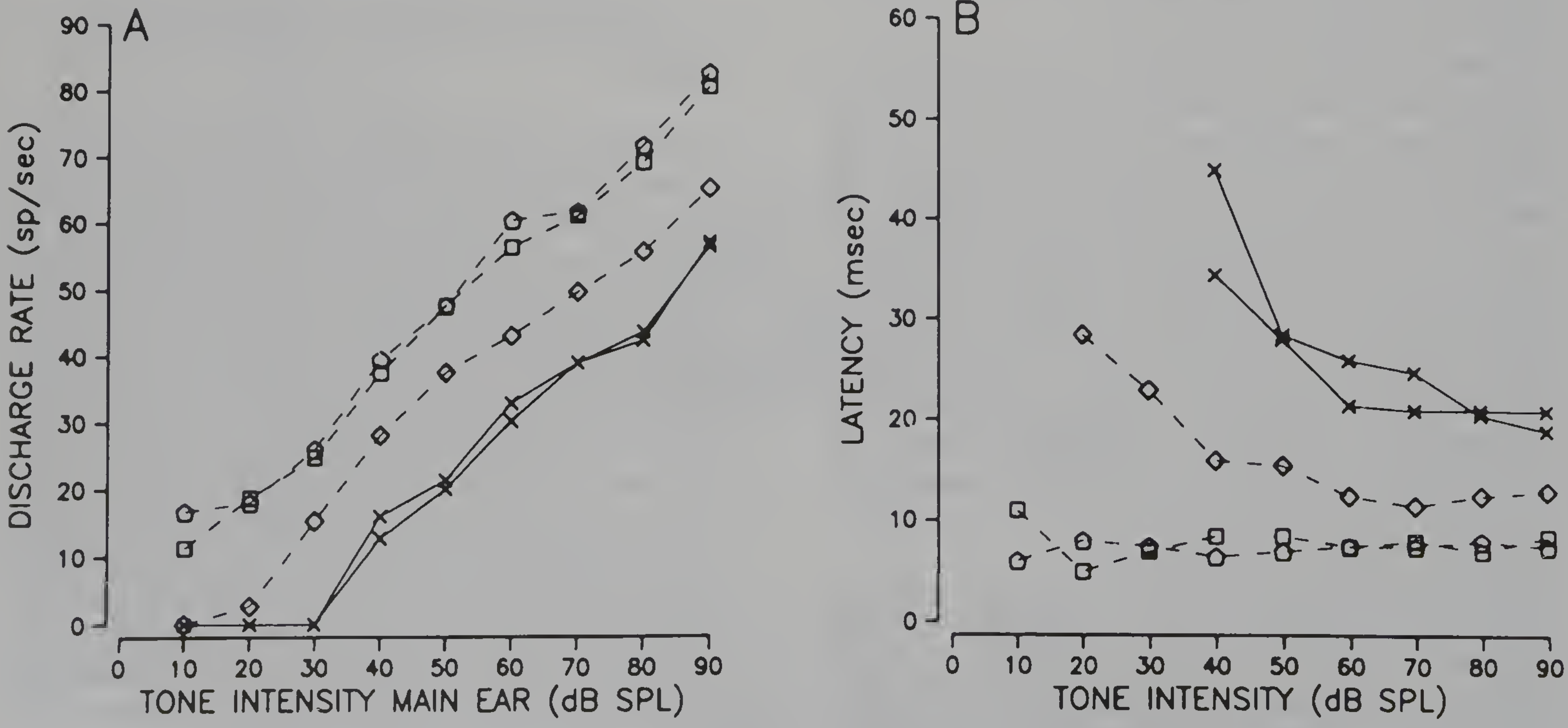


Fig. 6 (*a*) Rate-level and (*b*) latency-level characteristics of an ipsilateral type OCB fiber at CF without (solid lines) and with (dashed lines) background noise. (From Liberman, Ref. 57.)

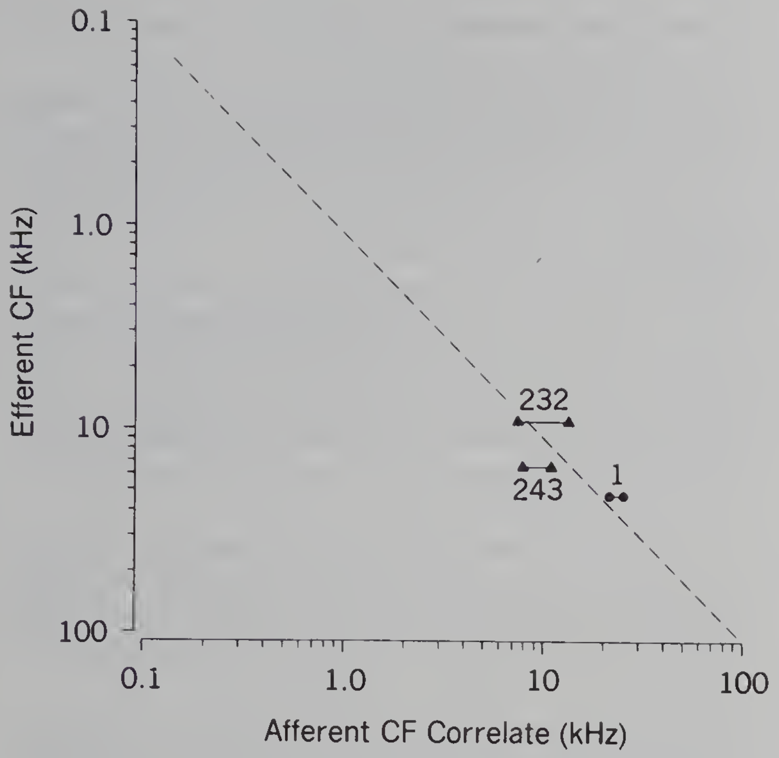


Fig. 7 Efferent CF vs. afferent CF correlate. (●) and (▲) represent ipsi and contra fibers, respectively. (From Ref. 62.)

fibers signal the level of a tone presented together with broadband masking noise.

Figure 7 shows the relationship between efferent fiber CF and the afferent CF correlate, that is, the CF of afferent fibers innervating the same cochlear place (data from Liberman and Brown[62]). Three efferent fibers are represented in the figure as three short line segments. The results for these three efferent fibers indicate that the efferent CF and the afferent CF correlate are *approximately* equal. However, there are theoretical grounds for predicting that the efferent fiber CF should be *systematically lower* than the afferent CF correlate (Fig. 2, Kim[63]). This prediction is based on the hypotheses that the OHCs provide active biomechanical amplification of cochlear partition motion (Chapter 108) and that the circuit consisting of the OHCs, type II spiral ganglion neurons and medial OCB efferent neurons forms a distinct OHC subsystem.[60] This prediction states that the characteristic basal travel of an outer spiral fiber[64] should be reflected in the CF of a medial efferent fiber innervating OHCs relative to the afferent CF correlate; the latter represents the CF of an IHC afferent rather than an OHC afferent. For an outer spiral fiber, the distance along the cochlear partition between where it leaves the habenula perforata and the group of OHCs it contacts appears to be about 0.5–1.5 mm. According to the cat cochlear map,[2] this distance represents about 0.14–0.5 octave. Two of the three efferent fibers in Fig. 7 are consistent with this theoretical prediction in terms of both the direction of the shift and the amount of shift (i.e., the basal border of OHC innervation is about 0.25–0.88 octave away). Observations made in the guinea pig by Robertson and his colleagues[1,65] are also consistent with the above prediction. Currently, the number of efferent fibers with combined physiological and anatomical descriptions is rather small. Such data from a larger number of efferent fibers would allow us to evaluate the theoretical prediction about the efferent–afferent CF relationship more reliably.

Acknowledgment

This work was supported in part by NIDCD, NIH. We thank Dr. Joseph Hind, University of Wisconsin, for his comments and help in improving the clarity of the text.

REFERENCES

1. D. Robertson, "Horseradish Peroxidase Injection of Physiologically Characterized Afferent and Efferent Neurones in the Guinea Pig Spiral Ganglion," *Hear. Res.*, Vol. 15, 1984, pp. 113–121.
2. M. C. Liberman, "The Cochlear Frequency Map for Cat: Labeling Auditory-Nerve Fibers of Known Characteristic Frequency," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 1441–1449.
3. N. Y. S. Kiang, T. Watanabe, E. C. Thomas, and L. F. Clark, *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*, MIT Press, Cambridge, MA, 1965.
4. M. C. Liberman, "Auditory-Nerve Response from Cats Raised in Low-Noise Chamber," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 442–455.
5. D. O. Kim and C. E. Molnar, "A Population Study of Auditory Nerve Fibers: Comparison of Spatial Distributions of Average-Rate and Phase-Locking Measures of Responses to Single Tones," *J. Neurophysiol.*, Vol. 42, 1979, pp. 16–30.
6. D. O. Kim, K. Parham, S. O. Chang, and J. G. Sirianni, "Spatial Response Profiles of Posteroventral Cochlear Nucleus Neurons and Auditory-Nerve Fibers in Unanesthetized Decerebrate Cats: Response to Pure Tones," *J. Acoust. Soc. Am.*, Vol. 86, 1991, pp. 2804–2817.
7. M. B. Sachs, C. C. Blackburn, and E. D. Young, "Rate-Place and Temporal-Place Representations of Vowels in the Auditory Nerve and Anteroventral Cochlear Nucleus," *J. Phonetics*, Vol. 16, 1988, pp. 37–53.
8. M. B. Sachs and E. D. Young, "Encoding of Steady-State Vowels in the Auditory Nerve: Representation in Terms of Discharge Rate," *J. Acoust. Soc. Am.*, Vol. 66, 1979, pp. 470–479.
9. R. R. Pfeiffer and D. O. Kim, "Response Patterns of Single Cochlear Nerve Fibers to Click Stimuli: Descriptions for Cat," *J. Acoust. Soc. Am.*, Vol. 52, 1975, pp. 1669–1677.
10. E. F. Evans, "The Dynamic Range Problem: Place

and Time Coding at the Level of Cochlear Nerve and Nucleus," in J. Syka and L. Aitkin (Eds.), *Neural Mechanisms of Hearing*, Plenum Press, New York 1981, pp. 69–84.

11. J. A. Costalupes, "Representation of Tones in Noise in the Responses of Auditory Nerve Fibers in Cats," *J. Neurosci.*, Vol. 5, 1985, pp. 3261–3269.
12. W. P. Shofner and M. B. Sachs, "Representation of a Low-Frequency Tone in the Discharge Rate of Populations of Auditory Nerve Fibers," *Hear. Res.*, Vol. 21, 1986, pp. 91–95.
13. M. I. Miller, P. E. Barta, and M. B. Sachs, "Strategies for the Representation of a Tone in Background Noise in the Temporal Aspects of the Discharge Patterns of Auditory-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 655–679.
14. D. O. Kim and K. Parham, "Auditory Nerve Spatial Encoding of High-Frequency Pure Tones: Population Response Profiles Derived from d' Measure Associated with Nearby Places along the Cochlea," *Hear. Res.*, Vol. 52, 1991, pp. 167–180.
15. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Wiley, New York 1966.
16. D. N. Elliot, L. Stein, and M. J. Harrison, "Determination of Absolute-Intensity Thresholds and Frequency Difference Thresholds in Cats," *J. Acoust. Soc. Am.*, Vol. 32, 1960, pp. 380–384.
17. J. E. Rose, J. F. Brugge, D. J. Anderson, and J. E. Hind, "Phase-Locked Response to Low-Frequency Tones in Single Auditory Nerve Fibers of Squirrel Monkey," *J. Neurophysiol.*, Vol. 30, 1967, pp. 769–793.
18. D. H. Johnson, "The Relationship of Post-stimulus Time and Interval Histograms to the Timing Characteristics of Spike Trains," *Biophysical J.*, Vol. 22, 1978, pp. 413–430.
19. M. C. Liberman, "Central Projections of Auditory-Nerve Fibers of Differing Spontaneous Rate, I: Anteroventral Cochlear Nucleus," *J. Comp. Neurol.*, 313, 1991, pp. 240–258.
20. M. B. Sachs and P. J. Abbas, "Rate Versus Level Functions for Auditory-Nerve Fibers in Cats: Tone-Burst Stimuli," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1835–1847.
21. E. F. Evans and A. R. Palmer, "Relationship between the Dynamic Range of Cochlear Nerve Fibers and Their Spontaneous Activity," *Exp. Brain Res.*, Vol. 40, 1980, pp. 115–118.
22. I. M. Winter, D. Robertson, and G. K. Yates, "Diversity of Characteristic Frequency Rate-Intensity Functions in Guinea Pig Auditory Nerve Fibers," *Hear. Res.*, Vol. 45, 1990, pp. 191–202.
23. M. C. Teich and S. K. Khanna, "Pulse-Number Distribution for the Neural Spike Train in the Cat's Auditory Nerve," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 1110–1128.
24. B. Delgutte, "Peripheral Auditory Processing of Speech Information: Implications from a Physiological Study of Intensity Discrimination," in M. E. H. Schouten (Eds.), *The Psychophysics of Speech Perception*, Martinus Nijhoff, Boston, 1987, pp. 333–353.
25. N. F. Viemeister, "Psychophysical Aspects of Auditory Intensity Coding," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function: Neurobiological Bases of Hearing*, Wiley & Sons, New York 1988, pp. 213–241.
26. E. D. Young and P. E. Barta, "Rate Response of Auditory Nerve Fibers to Tones in Noise Near Masked Threshold," *J. Acoust Soc. Am.*, Vol. 79, 1986, pp. 426–442.
27. D. O. Kim, S. O. Chang, and J. G. Sirianni, "A Population Study of Auditory-Nerve Fibers in Unanesthetized Decerebrate Cats: Response to Pure Tones," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 1648–1655.
28. W. P. Shofner and R. H. Dye, "Statistical and Receiver Operating Characteristics Analysis of Empirical Spike-Count Distributions: Quantifying the Ability of Cochlear Nucleus Units to Signal Intensity Changes," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 2172–2184.
29. M. C. Liberman, "Physiology of Cochlear Efferent and Afferent Neurons: Direct Comparisons in the Same Animal," *Hear. Res.*, Vol. 34, 1988, pp. 179–192.
30. R. A. Schmiedt, "Spontaneous Rates, Thresholds and Tuning of Auditory-Nerve Fibers in the Gerbil: Comparisons to Cat Data," *Hear. Res.*, Vol. 42, 1989, pp. 23–36.
31. C. D. Geisler, L. Deng, and S. R. Greenberg, "Thresholds for Primary Auditory Fibers Using Statistically Defined Criteria," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 1102–1109.
32. E. M. Relkin and D. G. Pelli, "Probe Tone Thresholds in the Auditory Nerve Measured by Two-Interval Forced-Choice Procedures," *J. Acoust. Soc. Am.*, Vol. 82, 1987, pp. 1679–1691.
33. R. L. Smith and J. J. Zwislocki, "Short-Term Adaptation and Incremental Responses of Single Auditory-Nerve Fibers," *Biol. Cybern.*, Vol. 17, 1975, pp. 169–182.
34. D. M. Harris and P. Dallos, "Forward Masking of Auditory Nerve Fiber Responses," *J. Neurophysiol.*, Vol. 42, 1979, pp. 1083–1107.
35. L. A. Westerman and R. L. Smith, "Rapid and Short-Term Adaptation in Auditory Nerve Responses," *Hear. Res.*, Vol. 15, 1984, pp. 249–260.
36. E. M. Relkin and J. R. Doucet, "Low-Spontaneous Rate Auditory Neurons Compared to High Spontaneous-Rate Neurons: I. Differences in Recovery from Prior Stimulation," *Assoc. Res. Otolaryngol. Abstr.*, Vol. 14, 1991, p. 127.
37. F.-G. Zeng, C. W. Turner, and E. M. Relkin, "Recovery from Prior Stimulation II: Effects upon Intensity Discrimination," *Hear. Res.*, 55, 1991, pp. 223–230.
38. J. L. Goldstein and N. Y. S. Kiang, "Neural Correlates of the Aural Combination Tone $2f_1 - f_2$," *Proc. IEEE*, Vol. 56, 1968, pp. 981–992.
39. D. O. Kim, "Cochlear Mechanics: Implications of Elec-

trophysiological and Acoustical Observations," *Hear. Res.*, Vol. 2, 1980, pp. 297–317.

40. J. H. Siegel, D. O. Kim, and C. Molnar, "Effects of Altering Organ of Corti on Cochlear Distortion Products $f_2 - f_1$ and $2f_1 - f_2$," *J. Neurophysiol.*, Vol. 47, 1982, pp. 303–328.
41. J. L. Hall, "Two-Tone Distortion Products in a Nonlinear Model of the Basilar Membrane," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1818–1828.
42. D. O. Kim, C. E. Molnar, and J. W. Matthews, "Cochlear Mechanics: Nonlinear Behavior in Two-Tone Responses as Reflected in Cochlear-Nerve-Fiber Responses and in Ear-Canal Sound Pressure," *J. Acoust. Soc. Am.*, Vol. 67, 1980, pp. 1704–1721.
43. J. L. Goldstein, "Auditory Nonlinearity," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 676–689.
44. M. B. Sachs and N. Y. S. Kiang, "Two-Tone Inhibition in Auditory-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 1120–1128.
45. E. Javel, "Suppression of Auditory-Nerve Responses I: Temporal Analysis, Intensity Effects and Suppression Contours," *J. Acoust. Soc. Am.*, Vol. 68, 1981, pp. 1735–1745.
46. N. M. Blachman, "Band-Pass Nonlinearities," *IRE Trans. Inform. Theory*, Vol. IT-10, 1964, pp. 162–164.
47. D. D. Greenwood, "What Is 'Synchrony Suppression'?" *J. Acoust. Soc. Am.*, Vol. 79, 1986, pp. 1857–1872.
48. R. R. Pfeiffer, "A Model for Two-Tone Inhibition in Single Cochlear-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 48, 1970, pp. 1373–1378.
49. W. S. Rhode, "Some Observations on Two-Tone Interaction Measured with the Mossbauer Effect," in E. F. Evans and J. P. Wilson (Eds.), *Psychophysics and Physiology of Hearing*, Academic Press, London 1977, pp. 27–38.
50. L. Robles, M. A. Ruggero, and N. C. Rich, "Basilar Membrane Mechanics at the Base of the Chinchilla Cochlea. I. Input-Output Functions, Tuning Curves, and Response Phases," *J. Acoust. Soc. Am.*, Vol. 80, 1986, pp. 1364–1374.
51. C. D. Geisler, G. K. Yates, R. B. Patuzzi, and B. M. Johnstone, "Saturation of Outer Hair Cell Receptor Currents Causes Two-Tone Suppression," *Hear. Res.*, Vol. 44, 1990, pp. 241–256.
52. P. M. Sellick and I. J. Russell, "Two-Tone Suppression in Cochlear Hair Cells," *Hear. Res.*, Vol. 1, 1979, pp. 227–236.
53. D. O. Kim, "A Review of Nonlinear and Active Cochlear Models," in J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Eds.), *Peripheral Auditory Mechanisms*, Springer Verlag, Berlin, 1986, pp. 239–247.
54. M. B. Sachs and E. D. Young, "Encoding of Steady-State Vowels in the Auditory Nerve: Representation in Terms of Discharge Rate," *J. Acoust. Soc. Am.*, 66, 1979, pp. 470–479.
55. E. D. Young and M. B. Sachs, "Representation of Steady-State Vowels in the Temporal Aspects of the Discharge Patterns of Populations of Auditory Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 66, 1979, pp. 1381–1403.
56. J. Fex, "Auditory Activity in Centrifugal and Centripetal Cochlear Fibers in Cat," *Acta Physiol. Scand.*, Vol. 64, 1962, 43–57.
57. M. C. Liberman, "Response Properties of Cochlear Efferent Neurons: Monaural vs. Binaural Stimulation and the Effects of Noise," *J. Neurophysiol.*, Vol. 60, 1988, pp. 1779–1798.
58. R. Galambos, "Suppression of Auditory Activity by Stimulation of Efferent Fibers to Cochlea," *J. Neurophysiol.*, Vol. 19, 1956, pp. 424–437.
59. M. L. Wiederhold, "Variations in the Effects of Electrical Stimulation of the Crossed Olivocochlear Bundle on Cat Auditory Nerve Fiber Responses to Tone Bursts," *J. Acoust. Soc. Am.*, Vol. 48, 1970, pp. 966–977.
60. D. O. Kim, "Functional Roles of the Inner- and Outer-Hair-Cell Subsystems in the Cochlea and Brainstem," in C. I. Berlin (Ed.), *Hearing Science*, College-Hill Press, San Diego, 1984, pp. 241–262.
61. R. L. Winslow and M. B. Sachs, "Effects of Electrical Stimulation of the Crossed Olivocochlear Bundle on Auditory Nerve Response to Tones in Noise," *J. Neurophysiol.*, Vol. 57, 1987, pp. 1002–1021.
62. M. C. Liberman and M. C. Brown, "Physiology and Anatomy of Single Olivocochlear Neurons in the Cat," *Hear. Res.*, Vol. 24, 1986, pp. 17–36.
63. D. O. Kim, "Active and Nonlinear Cochlear Biomechanics and the Role of Outer-Hair-Cell Subsystem in the Mammalian Auditory System," *Hear. Res.*, Vol. 22, 1986, pp. 105–114.
64. H. Spoendlin, "Innervation Patterns in the Organ of Corti of the Cat," *Acta Otolaryn.*, Vol. 67, 1969, pp. 239–254.
65. D. Robertson and M. Gummer, "Physiological and Morphological Characterization of Efferent Neurons in the Guinea Pig Cochlea," *Hear. Res.*, Vol. 20, 1985, pp. 63–77.

110

ANATOMY OF THE CENTRAL AUDITORY NERVOUS SYSTEM

DOUGLAS L. OLIVER

1 INTRODUCTION

The central auditory system is the neural system devoted to processing information about acoustic stimuli. It includes many design features typical of complex neural systems. For example, the auditory system is composed of specialized parts with specific functions, and it uses parallel processing to concurrently process acoustic information. Auditory pathways have a frequency (tonotopic) organization. The ascending auditory pathways (Fig. 1) are the routes auditory information follows as it enters the central nervous system. These ascending pathways emerge from the cochlea in the ear and synapse in the cochlear nucleus, superior olivary complex, inferior colliculus, medial geniculate body, and finally end in the cerebral cortex. The descending auditory pathways (Fig. 2) follow the reverse course and create feedback loops that ultimately influence the function of the cochlea. Together, these ascending and descending systems underlie our ability to perceive sounds. Connections of the auditory system with other neural systems allow the organism to respond to sound and to incorporate acoustical events into learning and memory. Detailed presentations of the anatomy of the auditory system can be found in several recent books.[1–3] Readers also may want to consult general neuroanatomical and neuroscience texts.[4–6]

2 DESIGN FEATURES OF THE CENTRAL AUDITORY NERVOUS SYSTEM

2.1 Specialized Parts

The central auditory nervous system, a complex biological system created by evolution, contains many parts (see Fig. 1). The function of each part is determined by the types of neurons present, their structure, their biophysical properties, their inputs, and their outputs. While the parts of the auditory system are evident, the specific functions of each part are not as clear. All parts are auditory by virtue of the connections between the cochlea and the brain made by the eighth cranial nerve. However, some cell types and patterns of connections are unique to only one part of the system, and these give rise to unique computational properties. For example, binaural processing is only possible when information from the two ears is combined in a single neural structure such as the *superior olivary complex* (Fig. 1, SOC).

2.2 Parallel Processing

Neural systems employ parallel processing on a massive scale. Parallel processing allows many neural events to happen concurrently. Thus, the nervous system is superior to a digital computer that employs only serial processing and executes one instruction at a time. Parallel processing allows more than one aspect of an acoustic stimulus to be processed simultaneously.

Two types of anatomical arrangements subserve parallel processing in the auditory system. *Within a pathway*,

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

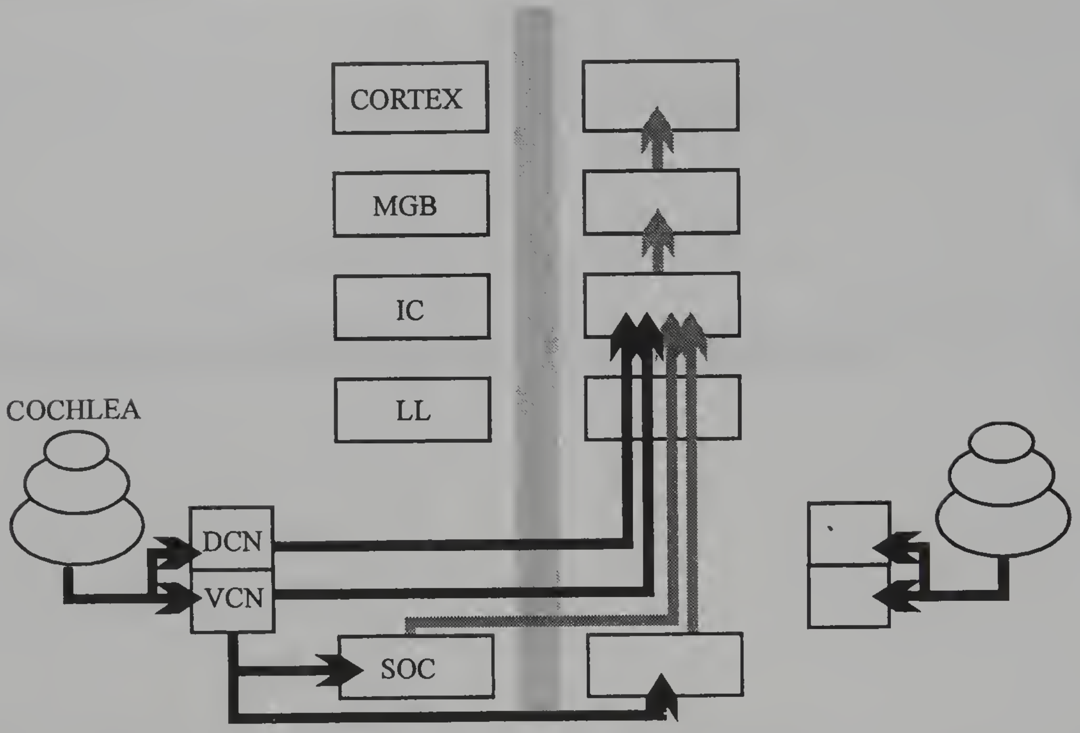


Fig. 1 Simplified wiring diagram of the ascending auditory pathways. The dominant flow of information from the ventral and dorsal cochlear nucleus (DCN, VCN) and superior olivary complex (SOC) crosses the midline (broad dotted line) to terminate in the inferior colliculus (IC). Pathways that convey monaural information (black) and binaural information (gray) both pass through the lateral lemniscus (LL). From the colliculus, projections ascend to the medial geniculate body (MGB) and auditory cortex (CORTEX).

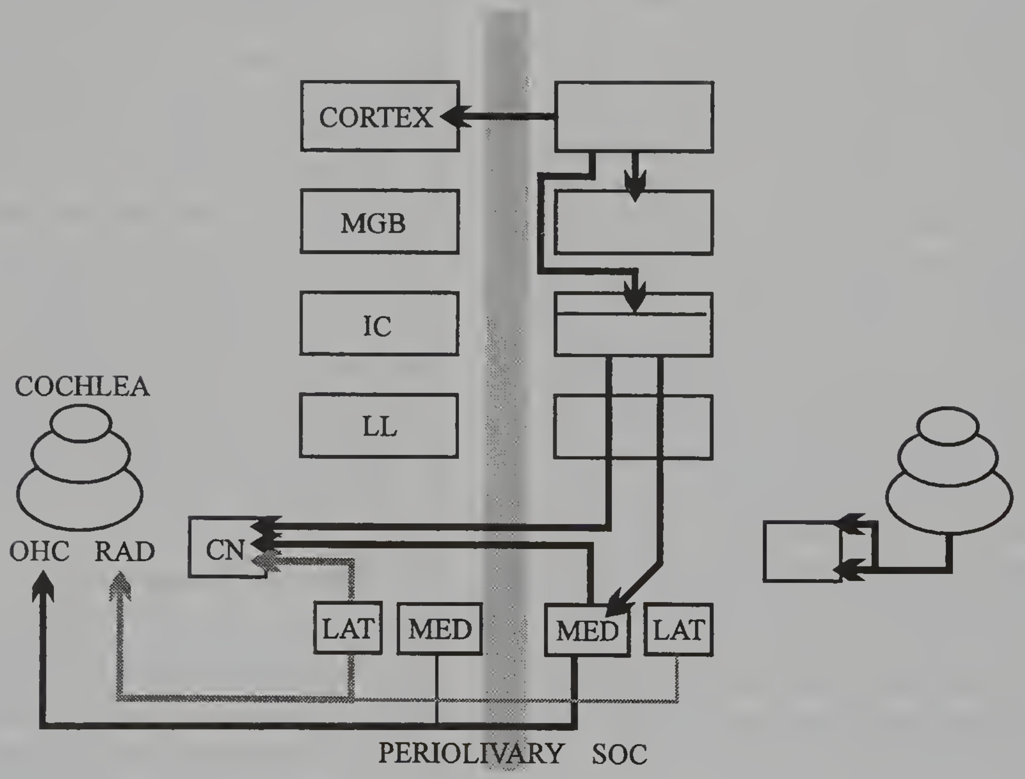


Fig. 2 Simplified wiring diagram of the descending auditory pathways. Descending pathways pass from the cortex (CORTEX) to the medial geniculate body (MGB) or inferior colliculus (IC). Descending pathways to the cochlea utilize the lateral (LAT) and medial (MED) periolivary nuclei of the superior olivary complex. The olivocochlear bundle terminates on the outer hair cells (OHC) and radial fibers (RAD) in the cochlea. The retrocochlear nucleus projections from MED and LAT terminate in the cochlear nucleus (CN).

there are hundreds or thousands of neurons of the same cell type. They have the same properties, the same inputs, and project to the same target but may differ in one variable such as frequency specificity. The inputs and outputs of these neurons are in parallel. *Parallel pathways* originate in different parts of the system and may function relatively independently. For example, the pathways that carry information from only one ear, the monaural pathways, may function in parallel to pathways that processes information about both ears, the binaural pathways.

2.3 Tonotopic Organization

The central auditory system has a precise frequency organization, for example, tonotopic organization, that is imparted by the connections with the cochlea. Frequency of the acoustic stimulus is encoded along the length of the cochlea (see Chapter 107), and this cochleotopic organization is transmitted to the brain. Thus, the frequency of the stimulus is translated to a spatial, tonotopic map in the brain. The tonotopic organization is maintained throughout the auditory pathways by focused, point-to-point connections that link the map at one level to the map at the next level.

3 ASCENDING AUDITORY PATHWAYS

The ascending auditory pathways begin with the eighth cranial nerve (vestibulo-cochlear) that connects the ear to the brain. Acoustic stimuli excite the hair cells of the cochlear partition, which activate the spiral ganglion and the eighth nerve. The peripheral portion of the nerve innervates the cochlea, and the central portion projects into the brain where it synapses in the *cochlear nucleus* of the medulla (Fig. 1, DCN, VCN, see also Chapter 109). Excitation of the nerve leads to excitation of the neurons in the cochlear nucleus. The cochlear nucleus sends projections centrally to the superior olivary complex in the medulla (Fig. 1, SOC) and the inferior colliculus in the midbrain (Fig. 1, IC). The superior olivary complex is the first site where information from both ears is combined, and it also projects to the inferior colliculus. From the inferior colliculus, the auditory pathway ascends to the medial geniculate body in the thalamus (Fig. 1, MGB) and the auditory cortex in the telencephalon (Fig. 1, CORTEX).

3.1 Cochlear Nucleus

The cochlear nucleus is the first processing center in the auditory pathway and is composed of dorsal and ventral nuclei (Fig. 1, DCN, VCN). The nerve branches into an ascending branch that terminates exclusively in the *anteroventral cochlear nucleus* (Fig. 3, AVCN) while the descending branch terminates in the *posteroventral cochlear nucleus* (Fig. 3, PVCN) and then continues into the *dorsal cochlear nucleus* (Fig. 3, DCN). The projections of the eighth nerve are tonotopically organized. The fibers from the basal, high-frequency part of the cochlea terminate in the dorso-caudal part of the anteroventral, posteroventral, and dorsal cochlear nuclei. The fibers from the apical, low-frequency part of the cochlea terminate in the rostro-ventral part of the same nuclei.

Unique cell types define each part of the cochlear nucleus, process information about acoustic stimuli, and project to different parts of the auditory system. In the anteroventral cochlear nucleus, the principal cell types are the bushy (Fig. 3, B, G) and stellate cells (Fig. 3, S). Bushy cells have very round cell bodies and usually only one proximal dendrite that forms a distinctive bushlike dendritic tree. There are two varieties of bushy cells, the spherical (Fig. 3, B) and the globular (Fig. 3, G). The spherical bushy cells receive a unique type of synaptic ending from the vestibulo-cochlear nerve, the endbulb of Held. Only one or two endbulbs terminate on each bushy cell, but these are sufficient in size to surround the cell body and provide massive excitation in response to stimulation of the nerve. The spherical bushy cells project to the superior olivary complex. The globular bushy cells are located more posteriorly in the anteroventral cochlea nucleus, receive synaptic inputs from more numerous and smaller axonal endings from the nerve, and also project to the superior olivary complex. Stellate cells (Fig. 3, S) are the second major cell type in the anteroventral cochlear nucleus. Unlike bushy cells, the stellate cells receive numerous small inputs from the eighth nerve primarily on their dendrites. Many stellate cells project to the inferior colliculus of the midbrain and may not synapse in the superior olive. Some stellate cells have axons that remain within the cochlear nucleus.

In the posteroventral cochlear nucleus, the octopus cell (Fig. 3, O), is usually the only cell type found in the most caudal part of that nucleus. This cell, which resembles its namesake in appearance, receives inputs from large numbers of eighth nerve fibers as they course through the nucleus. The octopus cells project primarily to the periolivary nuclei that surround the main nuclei of the superior olive and participate in feedback loops. They may also participate in ascending pathways through projections to the nuclei of the lateral lemniscus. Other cell types (not shown) are found in the rostral posteroventral cochlear nucleus and are similar to the bushy and stellate cells of the anteroventral cochlear nucleus.

In the dorsal cochlear nucleus, the principal cell type is the fusiform cell (Fig. 3, F). This cell has a vertically oriented dendritic tree that receives many inputs from the eighth nerve. The main target of the fusiform cell is the

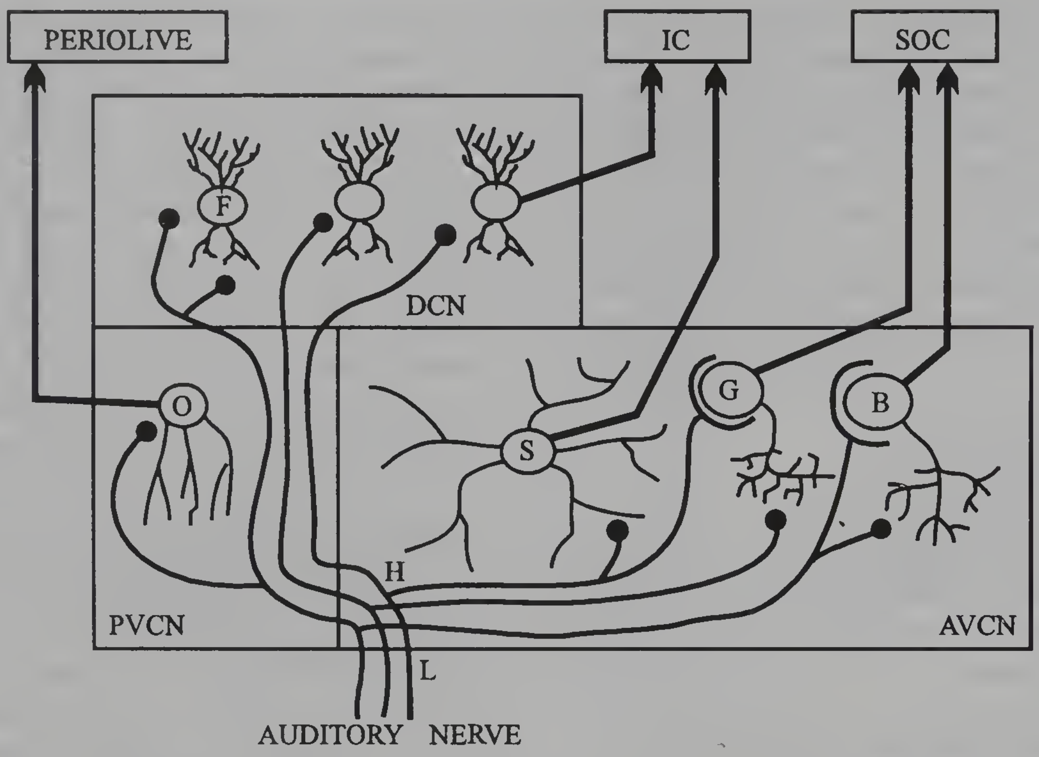


Fig. 3 Cell types and connections of the cochlear nucleus. In the anteroventral cochlear nucleus (AVCN), spherical bushy (B) and globular bushy cells (G) receive tonotopically organized inputs from the auditory nerve. Low frequency (L) is ventral and rostral while high frequencies (H) are caudal and dorsal. Bushy cells project to the superior olivary complex (SOC). Stellate cells (S) in AVCN and fusiform cells (F) in the dorsal cochlear nucleus (DCN) also receive inputs from auditory nerve but project to the inferior colliculus (IC). Octopus cells (O) in the posteroventral cochlear nucleus (PVCN) project to the periolivary SOC.

inferior colliculus. Most of the other cells (not shown) in the dorsal cochlear nucleus are smaller internuclear neurons except the giant cells whose axons project either to the midbrain or to the opposite cochlear nucleus.

3.2 Superior Olivary Complex

The superior olivary complex (Fig. 1, SOC, and Fig. 4) is the primary site of binaural interactions in the auditory system. As shown in Fig. 1, it receives inputs from the cochlear nucleus on the same and opposite side of the brain. The two principal nuclei of the olive are the *medial superior olive* (Fig. 4, MSO) and the *lateral superior olive* (Fig. 4, LSO). These nuclei project to the inferior colliculus (Fig. 5). Surrounding the principal nuclei are the *periolivary nuclei*. Some periolivary cells participate in ascending projections (see Section 2.2) while others are part of feedback loops (Fig. 2 and Section 4).

The medial superior olive (Fig. 4, MSO) is designed to encode interaural time differences. This nucleus has a three-dimensional structure that represents frequency in the dorsoventral direction and time in the rostrocaudal direction. Dorsal cells in the medial superior olive respond to low frequency (Fig. 4, L). Ventral cells respond to higher frequencies (Fig. 4, H). However, the highest frequencies are under represented compared to other auditory nuclei. The rostrocaudal dimension of the medial superior olive may encode interaural time differences (see Chapters 111 and 117). Each row of neurons receives excitatory inputs from the ventral cochlear nucleus bilaterally, for example, from both sides. These axons (Fig. 4, B) are from the spherical bushy cells and run along the rostrocaudal axis of the row. Specific locations along the horizontal azimuth (translated to interaural time differences) appear to be represented in different parts of the medial superior olive. Neurons in the rostral part of the nucleus respond to stimuli in front of the animal where there is zero interaural time difference (Fig. 4, 0 ITD). Neurons in the caudal part respond best to larger time differences usually to stimuli from the contralateral side (Fig. 4, CONTRA ITD). In addition to the inputs from the cochlear nucleus, the medial superior olive also receives inputs from the medial nucleus of the trapezoid body (Fig. 4, MNTB). These may provide inhibitory inputs to the medial superior olive and decrease the responses to unfavorable interaural time differences.

The anatomy of the lateral superior olive may be specialized to encode interaural intensity (level) differences by means of an interesting inhibitory mechanism. This nucleus contains neurons arranged in an S-shaped layer with a complete tonotopic map. Low frequencies are lat-

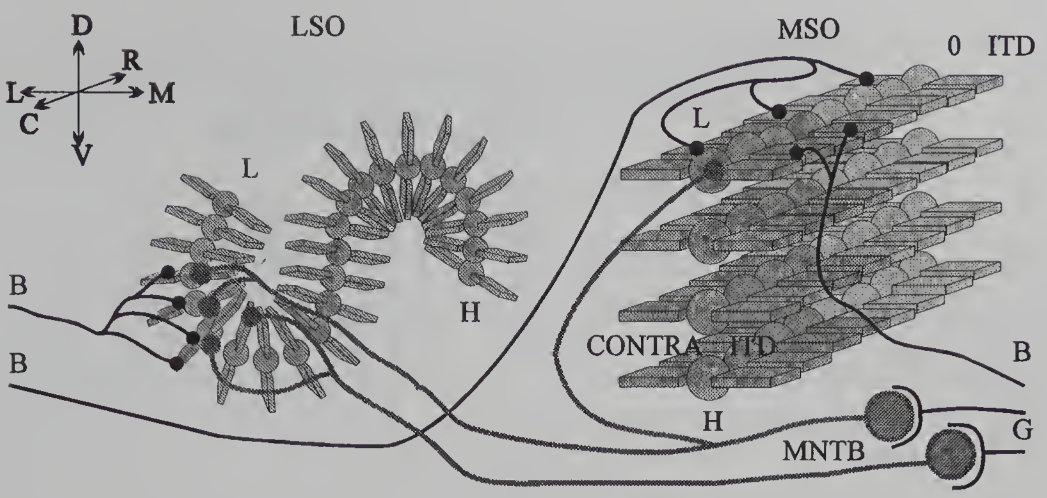


Fig. 4 Principal nuclei of the superior olivary complex. Medial superior olive (MSO) has bilateral inputs from axons of spherical bushy cells (B). Lateral superior olive (LSO) has similar inputs from the ipsilateral side, but contralateral inputs from axons of globular bushy cells (G) arrive via a synapse in the medial nucleus of the trapezoid body (MNTB). Frequency organization (H, high frequency; L, low frequency) is present in MSO and LSO. Interaural time delay (ITD) in MSO ranges from zero ITD (0 ITD) to increasingly larger ITDs for stimuli on the contralateral side (CONTRA ITD). Directional arrows: C, caudal; D, dorsal; L, lateral; M, medial; R, rostral; V, ventral.

4, H). It lacks the long rostrocaudal dimension of the medial olive. The lateral superior olive has two primary inputs. Sounds in one ear excite the cochlear nucleus and the lateral superior olive on the same side of the brain via axons from spherical bushy cells (Fig. 4, B). However, stimuli to the opposite ear produce inhibition in the lateral superior olive via a two-neuron circuit. Globular bushy cells from the opposite cochlear nucleus synapse in the medial nucleus of the trapezoid body with specialized axonal endings, the calyx of Held. This is the largest synaptic ending in the brain, and it activates the neurons in the medial nucleus of the trapezoid body that, in turn, project to and inhibit the neurons in the lateral superior olive. The combination of bilateral inhibitory and excitatory inputs creates a mechanism sensitive to small interaural intensity differences.

3.3 Nuclei of Lateral Lemniscus

The *lateral lemniscus* (Fig. 1, LL) contains the fibers from the cochlear nucleus and superior olivary complex that travel to the inferior colliculus. These fibers surround and penetrate through the nuclei of the lateral lemniscus. The *ventral nucleus of the lateral lemniscus* (Fig. 5, VLL) receives inputs primarily from the opposite cochlear nucleus and projects to the inferior colliculus on the same side. Neurons in the ventral lateral lemniscus may respond only to acoustic stimuli in the opposite ear. In contrast, the *dorsal nucleus of the lateral lemniscus* (Fig. 5, DLL) may participate in binaural functions. The dorsal nucleus receives most of the same inputs as the inferior colliculus (see below). Major inputs arise in the superior olive and the cochlear nuclei. The output of the dorsal nucleus is a bilateral projection to the inferior colliculus. While the ventral and dorsal nuclei of the lateral lemniscus are distinguished in their responses to monaural and binaural stimuli, both may provide inhibitory inputs to the inferior colliculus. Ascending inhibitory inputs to the inferior colliculus could modify the bandwidth or the temporal responses of neurons in the inferior colliculus.

3.4 Inferior Colliculus

The *inferior colliculus* integrates the ascending auditory pathways. It contains three major parts, a central nucleus (Figs. 5 and 6, IC-CEN), a cortex (Figs. 5 and 6, DC), and a paracentral collection of nuclei (Figs. 5 and 6, PARA). The *central nucleus of the inferior colliculus* contains layers of neurons arranged to make a single tonotopic map. Inputs from both monaural and binaural pathways terminate on this map, but the inputs do not converge completely. Thus, many functional properties found in the lower system are preserved by the projections to the colliculus. For example, tonotopically organized inputs from the lateral and medial superior olive (Fig. 5, MSO, LSO) and the dorsal nucleus of the lateral lemniscus (Fig. 5, DLL) may transmit different binaural properties. The projections from the dorsal and ventral cochlear nucleus (Fig. 5, DCN, VCN) convey monaural information. Some inputs, for example, the ventral nucleus of the lateral lemniscus, seem to have a poor

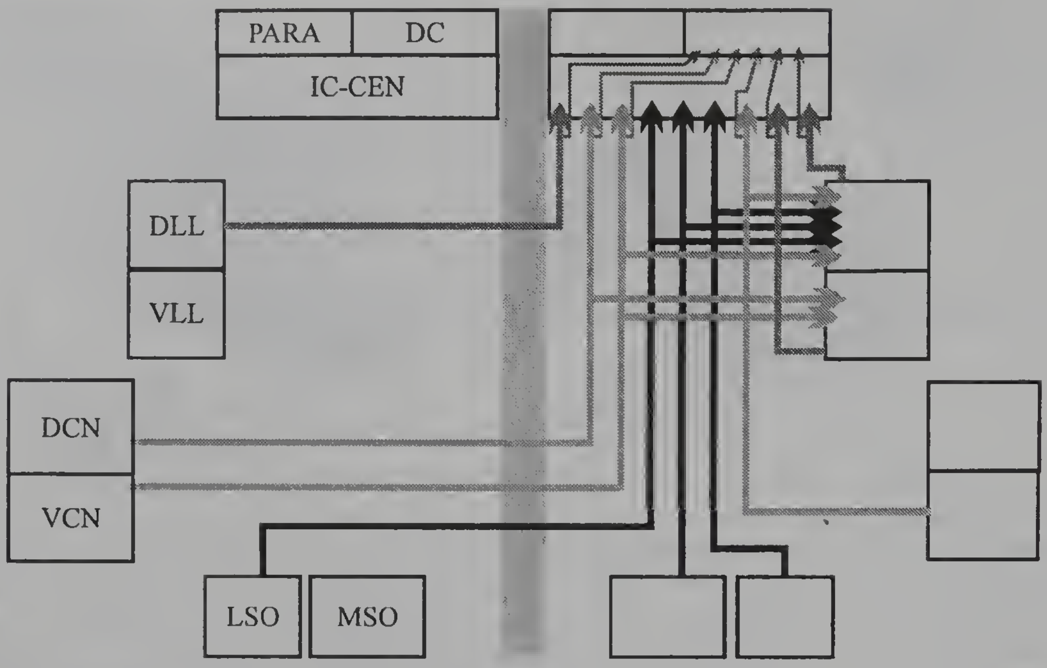


Fig. 5 Wiring diagram of parallel pathways to the inferior colliculus. Inferior colliculus contains central nucleus (IC-CEN), paracentral nuclei (PARA), and cortex (DC, dorsal cortex). Binaural pathways from superior olivary complex (LSO, MSO) are in black. Monaural pathways from cochlear nucleus (DCN, VCN) are in light gray. Inhibitory pathways from the nuclei of the lateral lemniscus (DLL, VLL) are in dark gray.

tonotopic organization but still terminate within the central nucleus.

Outside the central nucleus, the remaining inferior colliculus is less strictly concerned with the ascending auditory pathway and involves convergence with descending and other sensory information. The *dorsal cortex of the inferior colliculus* contains layers that are parallel to the surface of the inferior colliculus. The dorsal cortex (Fig. 5, DC) receives both ascending inputs and descending projections from the cerebral cortex. Its tonotopic organization is similar to that of the central nucleus and is primarily related to the organization of the ascending afferent fibers. *Paracentral nuclei of the inferior colliculus* receive minimal inputs from the ascend-

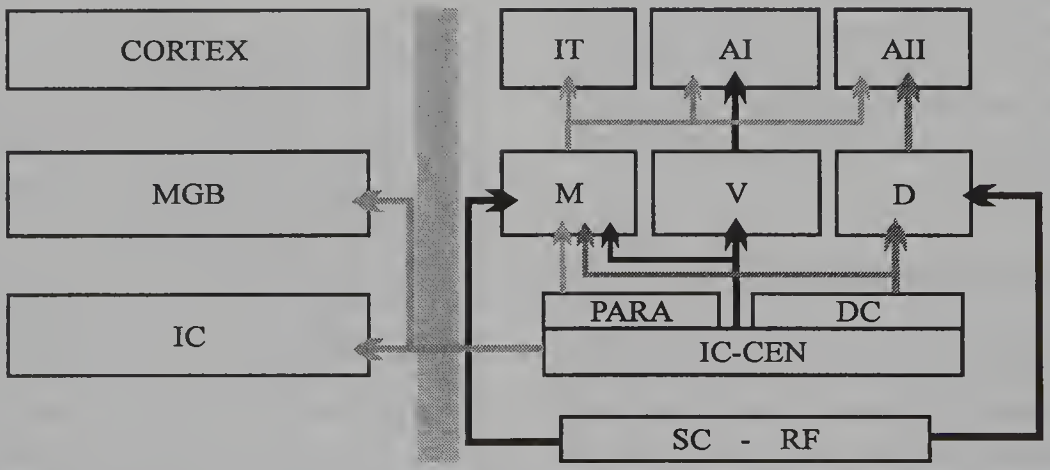


Fig. 6 Wiring diagram of the parallel pathways in the forebrain auditory system. Primary auditory cortex (CORTEX, AI) receives inputs from the central nucleus of the inferior colliculus (IC, IC-CEN) via a synapse in the ventral division (V) of the medial geniculate body (MGB). Multiple secondary cortical areas collectively illustrated as AII receive inputs from dorsal division (D) of the MGB whose inputs are from the dorsal cortex (DC) of the inferior colliculus and from the superior colliculus (SC) and reticular formation (RF). The medial division (M) of the MGB projects to all auditory cortical areas especially insular-temporal cortex (IT), and it receives inputs from all caudal structures including the paracentral nuclei (PARA) of the IC.

ing auditory pathways but do receive inputs from the spinal cord, somatosensory system, and reticular formation. There is little evidence for tonotopic organization within the paracentral nuclei.

3.5 Medial Geniculate Body

The *medial geniculate body* (Fig. 1, MGB) controls the flow of auditory information to the cerebral cortex. The geniculate contains ventral, dorsal, and medial divisions. The *ventral division of the medial geniculate body* (Fig. 6, V) contains layers that are tonotopically organized and functions as the primary relay of auditory information to cortex. Its main input is from the central nucleus of the inferior colliculus. The *dorsal division of the medial geniculate body* (Fig. 6, D) is more complex in anatomy and function. It contains several subdivisions and lacks a clear-cut tonotopic organization. Inputs are from outside the mainstream auditory system, for example, the dorsal cortex of the inferior colliculus, the superior colliculus (Fig. 6, SC), and the reticular formation (Fig. 6, RF) of the midbrain. The *medial division of the medial geniculate body* (Fig. 6, M) contains large cells and is often called the magnocellular nucleus. It receives converging inputs from auditory and nonauditory sources including spinal cord. The medial geniculate also may play a role in forming memories and emotional content through its projections to the amygdala in the limbic system.

3.6 Auditory Cortex

The ascending auditory pathways terminate in the *auditory cortex* (Figs. 1 and 6, CORTEX). In most mammals, including humans, the auditory cortex is recognized as having a primary cortex that is surrounded by secondary cortical regions. In humans, the *primary auditory cortex* is found in the temporal lobe in the first transverse gyrus of Heschl. The primary cortex (Fig. 6, AI) contains a complete tonotopic map and receives most of its inputs from the ventral division of the medial geniculate body. A number of *secondary auditory cortical areas* exist, for example, AII, insular, temporal (Fig. 6, AII, IT). Some areas like AII have well-developed tonotopic maps, others do not. The secondary areas usually receive inputs from both the dorsal division of the medial geniculate body and the primary cortex. One of the most elaborate auditory cortices is seen in the mustache bat. Specific cortical areas in the bat encode the constant-frequency and frequency-modulated components of the sounds emitted by the bat and the echoes of these sounds (see Chapter 113). In most mammals, there appear to be fewer cortical areas, and the specific functions of individual areas are less well defined. The primary connections of the auditory cortex are to the other parts of the cerebral cortex and telencephalon. These connections provide information about the acoustic environment to neural systems concerned with higher cognitive functions.

4 FEEDBACK LOOPS

The descending auditory pathways function to modify information traveling in the ascending pathways. They are separate from circuits that send information about sound to the motor systems or to other neural systems. There are two separate sets of descending inputs. One originates in the auditory cortex. The other system originates in the superior olive.

4.1 Descending Pathways from Cortex

Auditory cortex (Fig. 2) sends descending projections to the thalamus and midbrain. *Corticothalamic projections* terminate in the medial geniculate body (Fig. 2, MGB). In addition, *corticocollicular projections* synapse in the inferior colliculus (Fig. 2, IC), superior colliculus, and adjacent brainstem. Interestingly, the main component of the ascending pathways in the midbrain, the central nucleus of the inferior colliculus, receives the least descending input from the auditory cortex and the least feedback from cortex. Most of the descending inputs from telencephalon terminate in the dorsal cortex of the inferior colliculus or in the dorsomedial paracentral nucleus.

4.2 Olivocochlear System

Projections from the periolivary nuclei (Fig. 2) in the superior olive terminate in the cochlea and modify the response of the hair cells. The olivocochlear system is composed of medial and lateral parts that bear differing relationships to the hair cells of the cochlea. The lateral periolivary system (Fig. 2, LAT) projects primarily to the ipsilateral radial fibers (Fig. 2, RAD) that are the distal processes of the spiral ganglion beneath the inner hair cells in the cochlea. Depending on the species, the cells of origin for the *lateral olivocochlear system* are either within the lateral superior olive or in the adjacent dorsolateral periolivary nucleus. The *medial olivocochlear system* (Fig. 2, MED) terminates primarily on the outer hair cells (Fig. 2, OHC). Slightly more than half of the medial system terminates in the contralateral cochlea, and the remainder projects to the ipsilateral cochlea. Cells of origin for the medial system are in the medial periolivary nuclei including the medial nucleus of the trapezoid body, the ventral nucleus of the trapezoid body, and the dorsomedial periolivary nucleus.

Many of the same periolivary nuclei that contribute to the olivocochlear systems also send projections to the cochlear nucleus. These *retrocochlear nucleus projec-*

tions originate in the ventral periolivary nuclei (the ventral and lateral nuclei of the trapezoid body). The lateral group projects primarily to the ipsilateral cochlear nucleus, while the more medially placed ventral nucleus projects to the contralateral cochlear nucleus. Although the periolivary nuclei contain neurons that project to the cochlea, cochlear nucleus, and inferior colliculus, separate neurons apparently project to each site.

Both the periolivary nuclei and the cochlear nucleus receive descending inputs from the inferior colliculus (Fig. 2, IC). Most projections from the inferior colliculus to the superior olive terminate in the medial periolivary region (primarily the ventral nucleus of the trapezoid body). These projections arise principally in the central nucleus of the inferior colliculus. The central nucleus also projects directly to the cochlear nucleus where most of the descending fibers terminate in the dorsal cochlear nucleus. Neither the colliculo-olivary and colliculo-cochlear nucleus projections appear to be a direct continuation of the descending pathways from the auditory cortex.

REFERENCES

1. R. A. Altschuler, R. P. Bobbin, B. M. Clopton, and D. W. Hoffman (Eds.), *Neurobiology of Hearing: The Central Auditory System*, Raven Press, New York, 1991.
2. M. A. Merchán, J. M. Juiz, D. A. Godfrey, and E. Mugnaini (Eds.), *The Mammalian Cochlear Nuclei: Organization and Function. NATO ASI Series A*, Vol. 239, Plenum Press, New York, 1993.
3. D. B. Webster, A. N. Popper, and R. R. Fay (Eds.), *Springer Handbook of Auditory Research. The Mammalian Auditory Pathway: Neuroanatomy*, Vol. 1, Springer-Verlag, New York, 1992.
4. L. Heimer, *The Human Brain and Spinal Cord*, 2nd ed., Springer-Verlag, New York, 1995.
5. A. Brodal, *Neurological Anatomy in Relation to Clinical Medicine*, 3rd ed., Oxford University Press, New York, 1981.
6. E. R. Kandel, J. H. Schwartz, and T. M. Jessell (Eds.), *Principles of Neural Science*, 3rd ed., Appleton & Lange, Norwalk, CT, 1991.

111

ELECTROPHYSIOLOGY OF THE CENTRAL AUDITORY NERVOUS SYSTEM

TOM C. T. YIN AND PHILIP H. SMITH

1 INTRODUCTION

Acoustical information is transduced in the cochlea and presented to the central nervous system as trains of action potentials on several thousand auditory nerve fibers that converge on cells in the cochlear nucleus. From the cochlear nucleus to the auditory cortex are a series of nuclei that provide additional neural processing, the end result of which is the rich and complex perceptual experience that we recognize as hearing. In this chapter we will briefly review the electrophysiology of the central auditory system in an attempt to convey what is known about the neural processing by the brain. Because space limitations restrict the scope of this review, we will concentrate on topics about which more is known. Therefore we will discuss primarily processing in the auditory nuclei in the brainstem important for sound localization as derived from the responses of single cells, primarily from the cat. More extensive reviews are available elsewhere.[1,2].

1.1 Electrophysiological Techniques

Many techniques have been used to record physiological responses in the central auditory system. The placement and type of the recording electrode largely determines which neural element is being studied. Electrodes placed on the skull record the summed activity of many cells in the form of an evoked response, which is noninvasive and usable in the clinic on humans. To record the activity of single neurons, a metal or glass microelectrode (1–10 MΩ impedance) is advanced into the brain until action potentials, or "spikes," from neurons are encountered. When all of the spikes recorded at a given location are of the same size and shape, it is assumed that they originate from a single nearby neuron. With high-impedance (50–200 MΩ) glass microelectrodes, it is possible to penetrate individual cells and record from them intracellularly. This has the advantage that subthreshold synaptic events are also monitored, and if the microelectrode is also filled with a dye, the cell can be injected and visualized histologically.

It is sometimes advantageous to use in vitro techniques where neurons are removed from the brain and kept alive in a dish. This can be done by perfusing slices of brain with Ringer's solution or culturing dissociated neurons on an artificial nutrient substrate. In these artificial settings high-impedance or patch electrodes can stably record, from the whole cell or a membrane patch, membrane events that may be important in governing the cell's response.

1.2 General Properties of Central Auditory Structure

All of the information about the acoustical environment is conveyed to the central auditory system by auditory nerve fibers. This information is then relayed to brainstem nuclei before eventually reaching the cerebral cortex. We will first look at some general properties of central auditory structures that are applicable to all nuclei.

The most fundamental physiological characteristic of central auditory nuclei is the tonotopic organization that is set up in the cochlea and preserved at virtually every level of the auditory central nervous system (CNS). Like auditory nerve fibers, most cells are tuned to a narrow band of frequencies, and in each nucleus cells with sim-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

ilar characteristic frequencies (CFs) are grouped together with CFs changing progressively along one axis. Thus, beginning in the cochlear nucleus (CN) tonotopic organization is preserved in an orderly manner. Since the frequency axis is one dimensional but central auditory nuclei are two or three dimensional (some nuclei are flattened along one dimension), similar frequencies occupy a plane or strip, and a natural question is "What is encoded along the isofrequency strip?" There is some evidence that interaural parameters are encoded, but in most places the answer is not known.

The nearly uniform response properties of auditory nerve fibers (Chapter 109) contrasts sharply with the wide diversity of responses seen in central auditory cells. It is likely that this diversity in the CNS allows it to extract and encode particular features of the acoustic signal. A challenge for auditory physiologists is to understand both the mechanism by which this diversity arises as well as the functional meaning of the signals along each pathway. For most areas, we are still far from understanding these attributes.

The unique response features of any cell in the auditory CNS are the by-product of two basic parameters: its synaptic inputs and intrinsic membrane properties. The number, location, and sign (excitatory or inhibitory) of synaptic inputs determine the effectiveness of each input in regulating the ion channels, which control the flow of charged ions into or out of the cell. Channel activation or inactivation can be initiated by transmitter release from synapses, voltage changes across the membrane, or intracellular messengers. State changes in channels could depend on an auditory event or on the activation of inputs from other brain centers.

One of the major transformations in the CNS is the integration of inputs from the two ears. Studies of the sensitivity of central neurons to interaural time (ITDs) and level differences (ILDs) constitute a major part of our understanding of the mechanisms of central auditory processing and will be a focus of this review. The processing of many other important acoustic cues is not well understood.

In reviewing the pathways from the cochlea to the auditory cortex, the serial flow of information processing is obvious. Equally clear are paths for parallel processing, as certain pathways seem specialized for processing certain aspects of the acoustic input. For example, the coding of the two major binaural cues, ITDs and ILDs, use different pathways that diverge in the superior olivary complex.

Most experiments in this review were performed in anesthetized animals. Undoubtedly this affects the response properties of auditory neurons to varying degrees for different cells and at different sites in the CNS. Even in unanesthetized animals, the behavioral state of the animal plays a major role in determining the responses to certain sounds.

2 ELECTROPHYSIOLOGY OF CENTRAL AUDITORY SYSTEM

2.1 Cochlear Nucleus

The CN lies at the dorsolateral edge of the border between the pons and the medulla and is the termination site for all auditory nerve fibers. It has been subdivided into the anteroventral, posteroventral, dorsal cochlear nucleus. Although many cell types have only been defined anatomically, basic physiological features of a number of the principal cell types are available.

Bushy Cells Bushy cells, named from the appearance of their dendritic tree,[3] are situated in the anteroventral CN and are subdivided into spherical and globular bushy cells. Interruption of the output pathway for bushy cells results in a disruption of sound localization[4,5] and other auditory reflexes.[6] Synaptic inputs from auditory nerve fibers terminate primarily on the cell body in the form of very large specialized endings,[7–9] which are so large that a prepotential, thought to represent the presynaptic activity in the terminal, is sometimes seen when recording extracellularly.[10,11] Spherical bushy cells project bilaterally to the medial superior olive (MSO) and to the ipsilateral lateral superior olive (LSO) while globular cells project to the contralateral medial nucleus of the trapezoid body (MNTB) and bilaterally to many periolivary nuclei.[12–14]

A physiological classification scheme that shows good correlation with morphological classes Table 1) is based on the shape of the peristimulus time histogram in response to short tones at CF.[11,15] Spherical bushy cells with CFs > 1.5 kHz show peristimulus time histograms that resemble auditory nerve fiber responses and are called *primarylike* responses (Fig. 1). The histograms of globular bushy cells with high CFs can also be primarylike, but at higher stimulus intensities the first evoked spike and its refractory period are precisely timed, generating a peristimulus time histogram with a sharp peak followed by a brief depression and a resumption of sustained activity, which is called *primary-like-with-notch*.[16] Bushy cells with low best frequencies (<1 kHz) phase lock with higher temporal precision than auditory nerve fibers, indicating that they are not merely relay cells.[17]

Intracellular recordings from bushy cells in brain slices of the mouse reveal that the resistance of the membrane decreases during depolarization.[18,19] As a consequence the membrane time constant is smaller, allowing the cell to respond very rapidly to incoming synaptic

TABLE 1 Summary of Cochlear Nucleus Cell Types and Their Anatomical and Physiological Properties

Cell Type	Location	PSTH	Projection	Via	Membrane Characteristic
Spherical bushy	Ant. AVCN	Primarylike	Bilat. MSO Ipsi LSO	TB	Nonlinear
Globular bushy	Pos. AVCN Nerve root	Primary-like-with-notch	Contra MNTB, periolivary	TB	Nonlinear
Fusiform	DCN	Pauser Buildup	Contra IC	DAS	??
Multipolar (stellate)					
Type I	AVCN	Chopper	Contra IC	TB	Linear
Type II	PVCN	Onset chopper	Contra CN	IAS	?
Octopus	PVCN	Onset	DNLL, VNLL	IAS	?

Abbreviations: MSO, medial superior olive; LSO, lateral superior olive; MNTB, medial nucleus of the trapezoid body; DNLL, dorsal nucleus of the lateral lemniscus; VNLL, ventral nucleus of the lateral lemniscus; IC, inferior colliculus; TB, trapezoid body; DAS, dorsal acoustic stria; IAS, intermediate acoustic stria.

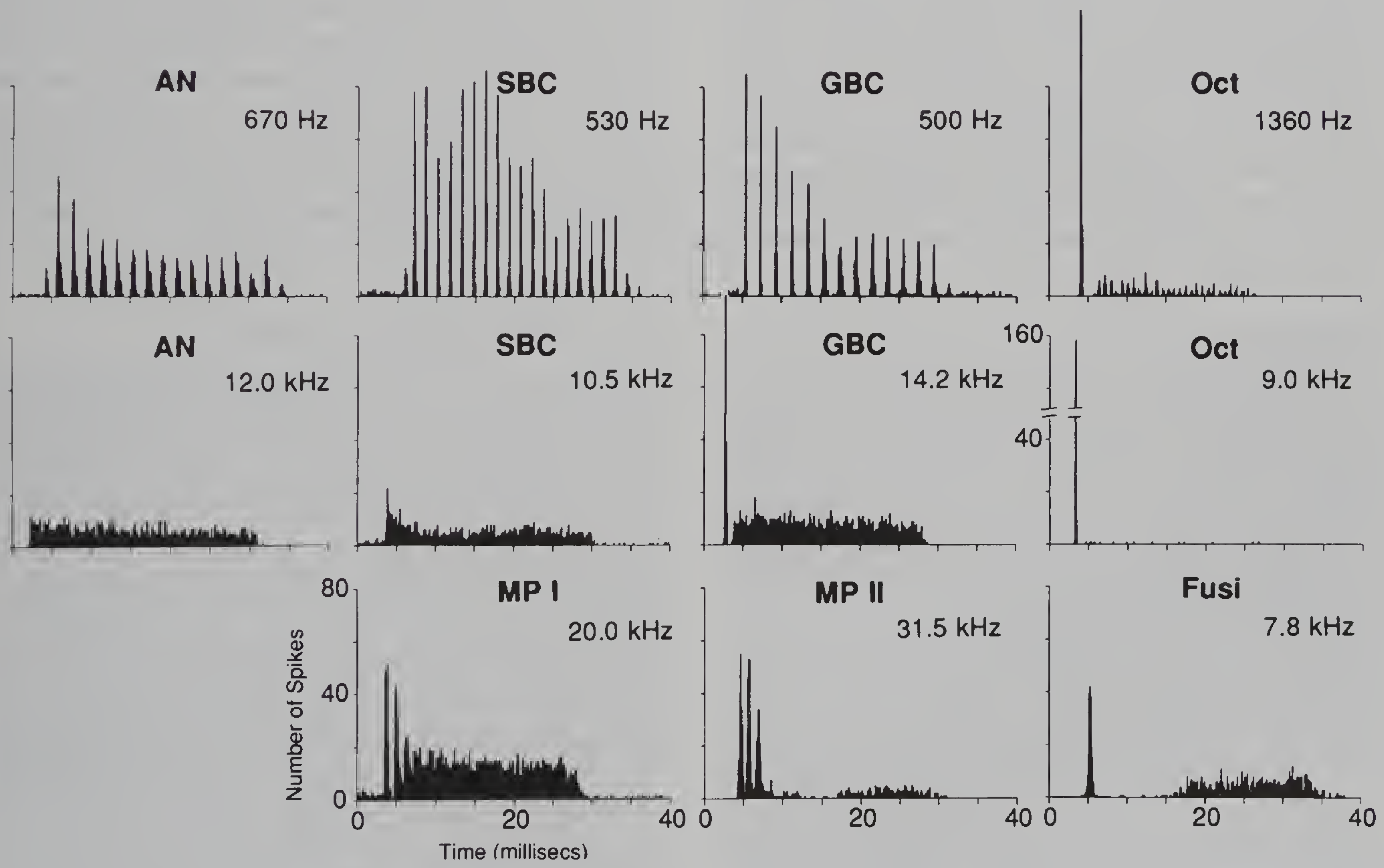


Fig. 1 Typical poststimulus time histograms of auditory nerve (AN, leftmost column) fibers and different cell types in the cochlear nucleus. Cells with low CFs that phase lock are in the top row; cells with high CFs are shown in the lower two rows. The cell types and their response histograms are: spherical bushy cell (SBC): primarylike; globular bushy cell (GBC): primary-like-with-notch; Octopus (Oct): onset; type I multipolar (MP I): chopper; type II multipolar (MP II): onset-chopper; fusiform (Fusi): pauser.

events, and subthreshold excitatory synaptic events can summate only over a narrow time window.

Fusiform and Giant Cells in the Dorsal Cochlear Nucleus Fusiform and giant cells are the principal projection neurons of the dorsal cochlear nucleus with axons leaving via the dorsal acoustic stria. The fusiform cells form a sheet a few hundred microns beneath the pial surface. Giant cells are scattered through the deep layers.[20,21] Fusiform cells project to the contralateral inferior colliculus (IC) while giant cells project to the IC or the contralateral CN.[22,23]

Compared to auditory nerve fibers, fusiform cells have higher thresholds and wider response areas, which can be multipeaked in frequency. Inhibitory inputs, active at and around CF, strongly influence responses to tones and noise.[24,26] The presence of these inhibitory sidebands allows for complex spectral processing, and the short-tone responses of fusiform cells are correspondingly more complex than bushy cells. They have been classified as pauser (onset response followed by a silent period then a gradual increase in discharge rate), buildup (pause response with no onset peak), or chopper (regular repetitive firing not related to the stimulus frequency).[15,27,28] Intracellular recordings from fusiform cells in brain slices of mice[29] or guinea pig[30] show that, if the membrane potential were set at different levels, depolarizing current steps can generate spike patterns resembling any of the three types described above, indicating that the intrinsic features of the cell membrane are capable of causing different response patterns.

Multipolar Cells Multipolar, or stellate, cells constitute a ubiquitous class of cells in the CN. With the electron microscope, two classes of multipolar cells have been distinguished in the anteroventral and posteroventral CN: type I cell bodies are sparsely covered with synaptic terminals, and their axons exit via the trapezoid body to project to the periolivary nuclei and IC; and type II cell bodies are densely innervated, and their axons travel via the intermediate acoustic stria.[31–33] Extracellular[11,34] and intracellular recordings[32,35] suggest that type I multipolar cells are associated with the chopper response. Multipolar cells in slices of the mouse show a linear current voltage characteristic and respond in a regular fashion when depolarized by intracellular current injection.[18,19] Type II multipolars show response patterns designated onset-choppers because of a short, but well-timed, chopperlike response at the onset of the histogram.[32] Unlike type I choppers, they respond over a much wider dynamic range (>90 dB), rarely show inhibitory sidebands, are much less regular during the sustained portion of a CF tone, and have axon terminals projecting to the opposite CN with pleomorphic vesicles (presumed inhibitory).

In addition, a distinct class of multipolar cells, termed octopus cells, are distinguished by their tentaclelike dendritic trees, are the sole occupant of the posterior pole of the posteroventral CN, and project to the lemniscal nuclei.[20] Extracellular and a few intracellular responses from the octopus cell area show that cells have broad tuning curves, an underrepresentation of low frequencies, and responses of the onset variety.[34]

Functional Transformations in the Cochlear Nucleus It should be clear from this discussion of the various cell types in the CN that the original incoming signal from the auditory nerve fibers, which show remarkably similar responses, are transformed in several complex ways. The output of the cochlear nucleus, as reflected in the response patterns of its cells (Fig. 1), appears to be specialized to encode particular aspects of the signal. For example, the onset cells seem to be well suited to encode the time of onset of the stimulus. Unfortunately, outside of the localization of sound, very little is known about the functional meanings of these transformations. For example, there are few behavioral consequences of a lesion of the output pathways from the dorsal cochlear nucleus.[36] Thus, we will concentrate the remainder of the review on circuits that encode ITDs and ILDs for sound localization, leaving open the vast perceptual attributes of pitch perception, loudness discrimination, speech perception, and so forth.

2.2 Superior Olivary Complex

The superior olivary complex consists of the LSO, MSO, MNTB, and a host of periolivary nuclei. Two of the three major pathways from both CN end in the superior olive. A major transformation is the integration of input from the cochlear nuclei of both sides, and hence from both ears. A large body of physiological, anatomical, and behavioral evidence[37] indicates that the two major interaural cues, ITD and ILD, are encoded by specialized cells in the MSO and LSO, respectively (Fig. 2). The classical duplex theory posits that ITDs are used to localize low-frequency sounds, where the CNS can encode temporal information, and ILDs for high-frequency sounds, where the shadowing effect of the head and body are effective (Chapter 117).

Encoding ITDs in the MSO Unfortunately, the MSO has proven to be a notoriously difficult nucleus to study electrophysiologically.[38] The segregation of afferents from the ipsilateral and contralateral anteroventral CN onto lateral and medial dendrites of MSO cells,

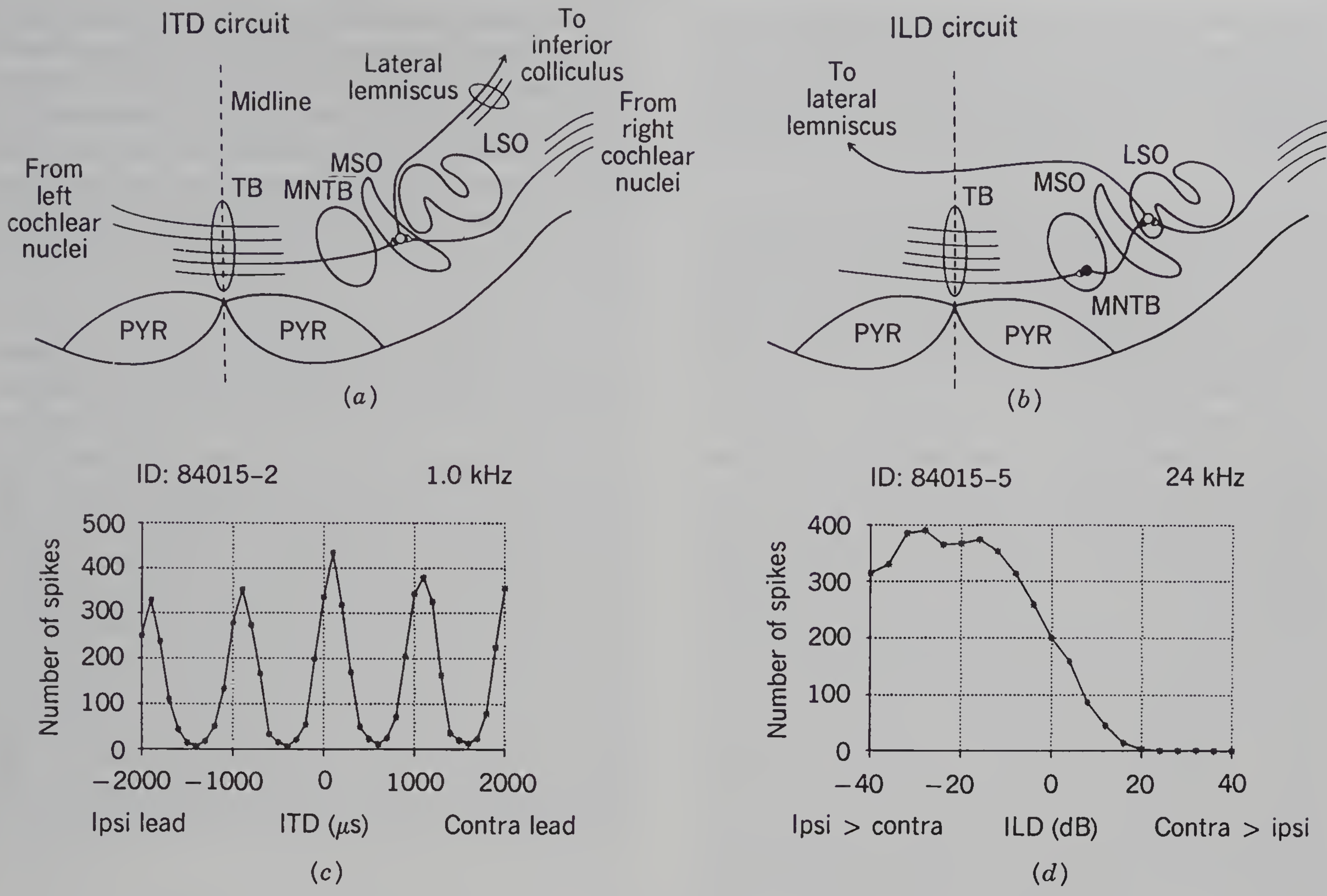

ITD circuit
Midline
Lateral lemniscus
To inferior colliculus
From right cochlear nuclei
From left cochlear nuclei
TB
MSO
MNTB
LSO
PYR
PYR
(a)
ILD circuit
To lateral lemniscus
TB
MSO
LSO
MNTB
PYR
PYR
(b)
ID: 84015-2
1.0 kHz
Number of spikes
500
400
300
200
100
0
−2000
−1000
0
1000
2000
Ipsi lead
ITD (μs)
Contra lead
(c)
ID: 84015-5
24 kHz
Number of spikes
400
300
200
100
0
−40
−20
0
20
40
Ipsi > contra
ILD (dB)
Contra > ipsi
(d)

Fig. 2 Schematic diagram of the two neural circuits and typical responses in the superior olive for encoding ITDs in the MSO (left) and ILDs in the LSO and MNTB (right). Inhibitory neurons are shown in black. The responses in (*c*) and (*d*) show sensitivity to ITDs and ILDs of CF tones in the MSO and LSO, respectively.

respectively,[39] suggests that binaural influences will play an important role. Indeed, from the earliest physiological studies of the MSO the importance of binaural cues has been emphasized.[40–42] Most, if not all, cells in the MSO are binaural and are sensitive to ITDs of pure tones[42–44] or broadband noise.[44]

Ironically, while there are few physiological studies of the MSO, there has been for some time a conceptual model[45] of how the nervous system might encode ITDs. The model assumes *first* that the inputs to the binaural cells carries timing information; *second*, that the binaural cells behave like coincidence detectors, that is, they respond most strongly when the inputs from both sides arrive simultaneously; and *third* that the inputs form neural delay lines along one axis of the nucleus. The result of these assumptions is that cells along that axis will be most sensitive to different ITDs in a graded fashion. This coincidence model has also been the key element in almost all psychophysical models of binaural processing (Chapter 117).

There is now evidence that all four conditions of the Jeffress model are valid for the MSO. First, as discussed above, the bushy cell inputs to the MSO are very highly phase locked.[17] Second, binaural cells respond maximally when the inputs from each side arrive simultaneously. The binaural response can be predicted by coincidence of the monaural response phases in the MSO.[42,44,46] Third, single afferent axons labeled with horseradish peroxidase to the contralateral, but not the ipsilateral, MSO have a delay line configuration, running from rostral to caudal.[47] Fourth, single-unit recordings indicate a map of ITDs across the rostral-caudal axis,[44] in accord with the caudally projecting delay lines from the contralateral afferents.

Both physiological and anatomical results indicate that each MSO contains a representation of the contralateral sound field. This representation is preserved at each of the higher stations, as would be expected from the anatomy: the MSO projects to the ipsilateral IC and nuclei of the lateral lemniscus. These physiological results are consonant with behavioral experiments in which unilateral lesions made at the level of the superior olivary complex or above result in deficits in localizing sounds in the contralateral but not the ipsilateral field.[5,48,49]

Encoding ILDs in the LSO While the MSO is important for encoding ITDs, the LSO and MNTB are important for encoding ILDs (Fig. 2). Most LSO cells are sensitive to ILDs and are excited by ipsilateral and inhibited by contralateral stimulation.[50,51] The ipsilateral excitation comes directly from spherical bushy cells of the anteroventral CN. The contralateral inhibition involves a disynaptic pathway originating from globular bushy cells that terminate on MNTB cells in the form of large calyces of Held.[14,52–54] Extracellular recordings in MNTB reveal large prepotentials presumed to reflect presynaptic current from the calyx,[55] and intracellular recordings from slices of the rat and mouse show large and fast excitatory postsynaptic potentials generated by the calyx.[56,57] MNTB cells project inhibitory glycinergic terminals to the ipsilateral LSO.[12,13,20,58] In the LSO electronmicrographic studies show terminals with round vesicles (presumed excitatory) on the distal dendrites and with pleomorphic vesicles (inhibitory) on cell bodies and proximal dendrites.[59] When recording from the LSO, the excitatory and inhibitory tuning curves are well matched.[50]

Since cells in the LSO are excited by stimulation of the ipsilateral ear and inhibited by stimulation of the contralateral (called IE in the standard convention of designating the contralateral followed by the ipsilateral response), they will respond to sound souces in the ipsilateral sound field, which contrasts with the EI response seen at all higher levels and the contralaterally responsive MSO cells. This apparent discrepancy is resolved by the projections of the LSO, which have a complex, but orderly, organization that has been termed the acoustic chiasm.[13,60] Cells in LSO project bilaterally with an asymmetry along the isofrequency axis. Most cells that project to the contralateral IC are from the high-frequency limb while most projecting ipsilaterally are from the low-frequency limb. Furthermore, most ipsilaterally projecting cells are glycinergic, as judged by staining of antibodies to glycine, and presumed inhibitory[61] while most contralaterally projecting cells are not glycinergic and presumed excitatory. The consequence of the crossing projection of the high-frequency LSO cells is that at all levels of the auditory system above the superior olive almost all high-frequency binaural cells that are sensitive to ILDs are EI and respond to stimuli presented in the contralateral sound field.

While all of the major nuclei in the superior olive are tonotopically organized, there is a nonuniform distribution of frequencies; both the MNTB and LSO have a predominance of high while the MSO is biased toward low frequencies.[38] These biases correspond to the functional roles of these nuclei (MSO for encoding ITDs and MNTB/LSO for ILDs) and the classical duplex theory outlined above. Comparative studies corroborate these functional differences as the ability of animals to detect ITDs is positively correlated with the size of the MSO.[62,63]

Encoding Both ITDs and ILDs of Complex Sounds It should be appreciated that the parallel processing scheme for ITDs and ILDs is a simplification. Psychophysical experiments have shown that the duplex theory holds strictly only for pure tones: humans can lateralize complex high-frequency signals using ITDs of the envelope.[64] Corresponding physiological studies have shown (1) that high-frequency auditory nerve fibers encode temporal information about the envelope of high-frequency amplitude modulated signals[65] and (2) that high-frequency LSO cells are sensitive to ITDs of the envelope of such waveforms.[66] In addition, the MSO does not encode ITDs exclusively, nor does the LSO encode ILDs exclusively. There are ILD-sensitive cells in the MSO,[41] and low-frequency LSO cells display ITD as well as ILD sensitivity.[66,67]

2.3 Inferior Colliculus and Nuclei of the Lateral Lemniscus

In the midbrain there is a convergence of the ascending auditory fibers into the nucleus of the lateral lemniscus and IC. There are only a few studies of the nuclei of the lateral lemniscus,[68,69] and we will describe this area only as it pertains to the IC. The IC is composed of a central nucleus (ICC) and a surrounding cortical shell, which appear functionally distinct. We will limit discussion to the ICC as it is the major integrative center for all ascending auditory pathways synapse in the ICC. In many ways our descriptions of the superior olive also apply to both the nuclei of the lateral lemniscus and IC: They are tonotopically organized and most cells are under binaural influence. As in the superior olivary complex, the low- and high-frequency binaural cells show sensitivity to ITDs and ILDs, respectively. Tonotopically, the cells in the ICC are arranged in sheets like the layers of an onion, with each layer representing an isofrequency lamina.[70,71]

ITD Sensitivity in the IC Most low-frequency cells in the ICC are binaural and sensitive to ITDs. These cells, like those in MSO, exhibit a *characteristic delay*, which is the ITD at which the delay curves for different frequency tones reach the same relative amplitude.[72, 73] ITD curves for noise and clicks are similar to the composite curve constructed from tonal responses by averaging all of the individual ITD curves at different frequencies.[74,75] In general the peaks of the composite curves occur at delays of the ipsilateral stimulus. The coincidence model was verified in its more general form of cross correlation by using noise stimuli correlated to different degrees or correlated only over specific frequency ranges.[76] The use of click stimuli also reveal the time course of the excitatory and inhibitory response components evoked on each side.[75]

ILD Sensitivity in the IC High-frequency cells in the ICC have been classified as monaural, EI, or EE.[72, 77, 78] Most ICC cells have sigmoidal, monotonic ILD functions, reflecting the EI responses while a minority have nonmonotonic, peaked ILD functions due to mixed inhibitory and facilitatory influences of the ipsilateral ear or nonmonotonic rate level functions of the contralateral input.[77] Some cells showed nonmonotonic functions with respect to both ILD and mean sound pressure level (SPL), as described in the gerbil ICC.[78] When signals with a time-varying envelope are used as inputs, these high-frequency cells also show sensitivity to ITDs of the envelope.[79]

Free-Field Studies in the ICC All of the above studies were done with the pinna removed so that stimuli could be delivered dichotically near the eardrum, which provides precise control over the acoustic stimulus, but circumvents the normal acoustic transformation provided by the outer ear.[80] Free-field stimulation also provides a more natural stimulus in which spatial receptive fields can be plotted directly, as demonstrated by studies in the barn owl.[81] A number of classification schemes have been used to describe the spatial receptive fields of cells in the ICC. Generally speaking, these receptive fields are consonant with the results from dichotic experiments. Omnidirectional cells respond similarly throughout the field, azimuth-selective cells have either a border or a peak response that remains stable with changes in SPL, and monaural cells increase their rate across the entire field as SPL increases.[82–84] Most azimuth-sensitive cells have receptive fields in the contralateral sound field.[82, 83] At low levels, the fields are dominated by the acoustical amplification by the pinna and most lie on its acoustic axis.[82, 84] Some azimuth-sensitive cells develop a second peak in the ipsilateral field at high levels, have low CFs, and likely correspond to the interaural phase-sensitive cells studied dichotically. Other azimuth-selective cells have sharp borders near the median plane and probably correspond to the EI cells, sensitive to ILDs.

2.4 Superior Colliculus

The superior colliculus (SC) is associated with the visual system and its function is thought to be visuomotor. However, some cells in the deep layers of the SC respond to acoustic as well as visual or somatic stimuli, and electrical stimulation evokes movements of the ears and head as well as the eyes. An interesting aspect of the SC is that cells in all three modalities form spatial maps that are in approximate register with one another. The details of the maps and their modifications during head, eye, and pinna movements are beyond the scope of this review.[2, 85]

2.5 Medial Geniculate Body

The ventral division of the MGB is the major thalamic relay for ascending auditory information. Much of what has been discussed above also pertains to responses in the MGB: Cells have sharp tuning curves, are tonotopically organized, binaurally influnced, and show sensitivity to ITDs and ILDs.[86, 87] The tonotopic organization of the MGB and its projection to the auditory cortex is complex.[88] A striking aspect of connectional studies of the MGB is the strong descending inputs from the auditory cortex as well as outputs from the MGB to subcortical areas such as the amygdala and striatum.[89] In the medial division, auditory responses are much more diffuse: Tuning curves and response areas are complex and wide with much variability and habituation.

2.6 Primary Auditory Cortex

In the cat the primary auditory cortex (AI) occupies the middle ectosylvian gyrus and both banks of the posterior ectosylvian sulcus and is surrounded by a number of secondary auditory areas that can be distinguished on cytoarchitecture,[90] thalamic inputs,[91] and physiological responses.[92] Physiologically these areas are differentiated by discontinuities in the tonotopic map across the surface as well as differences in the response properties of the neurons.[93] The result is multiple representations of the entire frequency spectrum so that each frequency is represented by a strip of cells in each of the fields, each strip running through the six layers of the neocortex.

What acoustic parameter, if any, is mapped along the isofrequency strips? In the high-frequency region most cells are binaural and can be classified as either EE or EI. These EE or EI cells appear to congregate in strips that run roughly orthogonal to the isofrequency strips.[94] With free-field stimuli azimuth-sensitive neurons are not randomly distributed across an isofrequency contour, but there is no clear point-to-point map.[95, 96]

Early studies of AI demonstrated cells with sensitivity to either ILDs or ITDs resembling those seen in lower centers.[97] Most high-frequency cells show ILD sensitivity; and low-frequency cells are sensitive to ITDs and appear to display characteristic delay.[98,99] Free-field studies have revealed omnidirectional, hemifield, and azimuth-selective units.[95,100] It is perhaps not surprising that responses seen in AI reflect inputs from lower centers. What is more interesting is the emergence of new properties that result from cortical mechanisms. The most dramatic examples of such mechanisms come from experiments in the auditory cortex of the mustached bat (Chapter 113).

Lesions of AI reveal behavioral deficits in sound localization. Unilateral lesions produce localization deficits in the contralateral, but not ipsilateral, sound field. If the lesion is confined to one isofrequency strip, the deficit is confined to the corresponding frequencies.[49] For other tasks, such as detection of sounds, frequency discrimination, or temporal pattern discrimination, results are mixed with some reports of severe deficits and others of little loss.[92] Interpretation of these results are complicated by differences in the behavioral task or training procedures.

3 FEEDBACK SYSTEMS TO THE AUDITORY PERIPHERY

While the primary flow of auditory information is toward the auditory cortex, each station along the pathway has cells that provide feedback onto lower centers. What modulatory effects these mechanisms might have on ascending auditory information is largely unknown. The most widely studied of the descending pathways are the two systems originating in the brainstem that project back to the auditory periphery: (1) motor neurons associated with the V and VII cranial nerves that innervate middle ear muscles and (2) the olivocochlear system innervating hair cells and/or nerve fibers of the cochlea.

In most animals, loud acoustic stimuli evoke a short latency reflexive contraction of the middle ear muscles bilaterally. In humans this reflex may only involve the stepedius.[101] There appears to be two basic types of stapedius motoneurons based on differences in cell location, axon course, and physiology.[102,103]

The olivocochlear system arises from cells in separate areas of the superior olive and is divided into a lateral olivocochlear system of unmyelinated axons, innervating auditory nerve processes beneath hair cells, and a medial system of myelinated axons, innervating outer hair cells of both the ipsilateral and contralateral cochlea.[104] The functional role of the olivocochlear system is still not understood. There are several popular hypotheses: It may serve to protect the ear from noise damage induced by loud sounds via the contralateral ear,[105] enhance detection of signals in noise,[106] or change the sensitivity of the cochlea.

REFERENCES

1. A. N. Popper and R. R. Fay, *The Mammalian Auditory Pathway: Neurophysiology*, Springer-Verlag, New York, 1992, pp. 1–431.
2. D. R. F. Irvine, "The Auditory Brainstem. A review of the Structure and Function of Auditory Brainstem Processing Mechanisms," in *Progress in Sensory Physiology*, D. Ottoson, (Ed.), Springer-Verlag, Berlin, 1986, pp. 1–277.
3. J. R. Brawer, D. K. Morest, and E. C. Kane, "The Neuronal Architecture of the Cochlear Nucleus of the Cat," *J. Comp. Neurol.*, Vol. 155, 1974, pp. 251–300.
4. C. N. Moore, J. H. Casseday, and W. D. Neff, "Sound Localization: The Role of the Commissural Pathways of the Auditory System of the Cat," *Brain Res.*, Vol. 82, 1974, pp. 13–26.
5. W. M. Jenkins and R. B. Masterton, "Sound Localization: Effects of Unilateral Lesions in Central Auditory Systems," *J. Neurophysiol.*, Vol. 47, 1982, pp. 987–1016.
6. M. Davis, D. S. Gendelman, M. D. Tischeler, and P. M. Gendelman, "A Primary Acoustic Startle Circuit: Lesion and Stimulation Studies," *J. Neurosci.*, Vol. 2, 1982, pp. 791–805.
7. R. Y. Cajal, *Histologie du systeme nerveux*, Vol. 1, Instituto Ramon Y Cajal, Madrid, Spain, 1909.
8. R. Lorente de No, "Anatomy of the Eighth Nerve. III. General Plan of the Structure of the Primary Cochlear Nuclei," *Laryngoscope*, Vol. 43, 1933, pp. 327–350.
9. J. R. Brawer and D. K. Morest, "Relations between Auditory Nerve Endings and Cell Types in the Cat's Anteroventral Cochlear Nucleus Seen with the Golgi Method and Nomarski Optics," *J. Comp. Neurol.*, Vol. 160, 1975, pp. 491–506.
10. R. R. Pfeiffer, "Anteroventral Cochlear Nucleus Wave Forms of Extracellularly Recorded Spike Potentials," *Science*, Vol. 154, 1966, pp. 667–668.
11. T. R. Bourk, "Electrical Responses of Neural Units in the Anteroventral Cochlear Nucleus of the Cat, Ph.D. Thesis, M.I.T., Cambridge, MA, 1976, pp. 1–385.
12. W. B. Warr, "Fiber Degeneration Following Lesions in the Anterior Ventral Cochlear Nucleus of the Cat," *Exp. Neurol.*, Vol. 14, 1966, pp. 453–474.
13. K. K. Glendenning, K. A. Hutson, R. J. Nudo, and R. B. Masterton, "Acoustic Chiasm II: Anatomical Basis of Binaurality in Lateral Superior Olive of Cat," *J. Comp. Neurol.*, Vol. 232, 1985, pp. 261–285.
14. P. H. Smith, P. X. Joris, L. H. Carney, and T. C. T. Yin, "Projections of Physiologically Characterized Globular Bushy Cell Axons from the Cochlear Nucleus of the Cat," *J. Comp. Neurol.*, Vol. 304, 1991, pp. 387–407.
15. R. R. Pfeiffer, "Classification of Response Patterns of Spike Discharge for Units in the Cochlear Nucleus: Tone

Burst Stimulation," *Exp. Brain Res.*, Vol. 1, 1966, pp. 220–235.

16. P. H. Smith and W. S. Rhode, "Characterization of HRP-labeled Globular Bushy Cells in the Cat Cochlear Nucleus," *J. Comp. Neurol.*, Vol. 266, 1987, pp. 360–376.
17. P. X. Joris L. H. Carney, P. H. Smith, and T. C. T. Yin, "Enhancement of Neural Synchronization in the Anteroventral Cochlear Nucleus. I: Responses to Tones at the Characteristic Frequency," *J. Neurophysiol.*, Vol. 71, 1994, pp. 1022–1036.
18. D. Oertel, "Synaptic Responses and Electrical Properties of Cells in Brain Slices of the Mouse Anteroventral cochlear Nucleus," *J. Neurosci.*, Vol. 3, 1983, pp. 2043–2053.
19. S. H. Wu and D. Oertel, "Intracellular Injection with Horeradish Peroxidase of Physiologically Characterized Stellate and Bushy Cells in Slices of Mouse Anteroventral Cochlear Nucleus," *J. Neurosci.*, Vol. 4, 1984, pp. 1577–1588.
20. K. K. Osen, "Cytoarchitecture of the Cochlear Nuclei in the Cat," *J. Comp. Neurol.*, Vol. 136, 1969, pp. 453–484.
21. K. K. Osen, "The Intrinsic Organization of the Cochlear Nuclei in the Cat," *Acta oto-laryngol.*, Vol. 67, 1969, pp. 352–359.
22. J. C. Adams and W. B. Warr, "Origins of Axons in the Cat's Acoustic Striae Determined by Injection of Horseradish Peroxidase into Severed Tracts," *J. Comp. Neurol.*, Vol. 170, 1976, pp. 107–122.
23. N. B. Cant and K. C. Gaston, "Pathways Connecting the Right and Left Cochlear Nuclei," *J. Comp. Neurol.*, Vol. 212, 1982, pp. 313–326.
24. E. F. Evans and P. G. Nelson, "The Responses of Single Neurons in the Cochlear Nucleus of the Cat as a Function of Their Location and Anesthetic State," *Exp. Brain Res.*, Vol. 17, 1973, pp. 402–427.
25. E. D. Young and W. E. Brownell, "Response to Tones and Noise of Single Cells in Dorsal Cochlear Nucleus of Unanesthetized Cats," *J. Neurophysiol.*, Vol. 39, 1976, pp. 282–300.
26. H. F. Voigt and E. D. Young, "Evidence of Inhibitory Interactions between Neurons in Dorsal Cochlear Nucleus," *J. Neurophysiol.*, Vol. 44, 1980, pp. 76–96.
27. W. S. Rhode, P. H. Smith, and D. Oertel, "Physiological Response Properties of Cells Labeled Intracellularly with Horseradish Peroxidase in Cat Dorsal Cochlear Nucleus," *J. Comp. Neurol.*, Vol. 213, 1983, pp. 426–447.
28. P. H. Smith and W. S. Rhode, "Electron Microscopic Features of Physiologically Characterized, HRP-labeled Fusiform Cells in the Cat Dorsal Cochlear Nucleus," *J. Comp. Neurol.*, Vol. 237, 1985, pp. 127–143.
29. J. A. Hirsch and D. Oertel, "Intrinsic Properties of Neurones in the Dorsal Cochlear Nucleus of Mice, In Vitro," *J. Phyiol. (London)*, Vol. 396, 1988, pp. 535–548.
30. P. B. Manis, "Membrane-Properties and Discharge Characteristics of Guinea Pig Dorsal Cochlear Nucleus Neurons Studied In Vitro," *J. Neurosci.*, Vol. 10, 1990, pp. 2338–2351.
31. N. B. Cant, "The Fine Structure of Two Types of Stellate Cells in the Anterior Division of the Anteroventral Cochlear Nucleus of the Cat," *Neuroscience*, Vol. 6, 1981, pp. 2643–2655.
32. P. H. Smith and W. S. Rhode, "Structural and Functional Properties Distinguish Two Types of Multipolar Cells in the Ventral Cochlear Nucleus," *J. Comp. Neurol.*, Vol. 282, 1989, pp. 595–616.
33. W. B. Warr, "Fiber Degeneration Following Lesions in the Multipolar and Globular Cell Areas in the Ventral Cochlear Nucleus of the Cat," *Brain Res.*, Vol. 40, 1972, pp. 247–270.
34. D. A. Godfrey, N. Y. S. Kiang, and B. E. Norris, "Single Unit Activity in the Dorsal Cochlear Nucleus of the Cat," *J. Comp. Neurol.*, Vol. 162, 1975, pp. 269–284.
35. W. S. Rhode, D. Oertel, and P. H. Smith, "Physiological Response Properties of Cells Labeled Intracellularly with Horseradish Peroxidase in Cat Ventral Cochlear Nucleus," *J. Comp. Neurol.*, Vol. 213, 1983, pp. 448–463.
36. R. B. Masterton and E. Moreland Granger, "Role of the Acoustic Striae in Hearing: Contribution of Dorsal and Intermediate Striae to Detection of Noises and Tones," *J. Neurophysiol.*, Vol. 60, 1988, pp. 1841–1860.
37. T. C. T. Yin and J. C. K. Chan, "Neural Mechanisms Underlying Interaural Time Sensitivity to Tones and Noise," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function: The Neurobiological Bases of Hearing*, New York, 1988, pp. 385–430.
38. J. J. Guinan, Jr., B. E. Norris, and S. S. Guinan, "Single Auditory Units in the Superior Olivary Complex. II: Location of Unit Categories and Tonotopic Organization," *Int. J. Neurosci.*, Vol. 4, 1972, pp. 147–166.
39. W. A. Stotler, "An Experimental Study of the Cells and Connections of the Superior Olivary Complex of the Cat," *J. Comp. Neurol.*, Vol. 98, 1953, pp. 401–432.
40. R. Galambos, J. Schwartzkopff, and A. Rupert, "Microelectrode Study of Superior Olivary Nuclei," *Am. J. Physiol.*, Vol. 197, 1959, pp. 527–536.
41. J. M. Goldberg and P. B. Brown, "Functional Organization of the Dog Superior Olivary Complex: An Anatomical and Electrophysiological Study," *J. Neurophysiol.*, Vol. 31, 1968, pp. 639–656.
42. J. M. Goldberg and P. B. Brown, "Response of Binaural Neurons of Dog Superior Olivary Complex to Dichotic Tonal Stimuli: Some Physiological Mechanisms of Sound Localization," *J. Neurophysiol.*, Vol. 32, 1969, pp. 613–636.
43. G. Moushegian, A. L. Rupert, and J. S. Gidda, "Functional Characteristics of Superior Olivary Neurons to Binaural Stimuli," *J. Neurophysiol.*, Vol. 38, 1975, pp. 1037–1048.
44. T. C. T. Yin and J. C. K. Chan, "Interaural Time Sensitivity in Medial Superior Olive of Cat," *J. Neurophysiol.*, Vol. 64, 1990, pp. 465–488.
45. L. A. Jeffress, "A Place Theory of Sound Localization," *J. Comp. Physiol. Psychol.*, Vol. 41, 1948, pp. 35–39.
46. G. Crow, A. L. Rupert, and G. Moushegian, "Phase Lock-

ing in Monaural and Binaural Medullary Neurons: Implications for Binarual Phenomena," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 493–501.

47. P. H. Smith, P. X. Joris, and T. C. T. Yin, "Projections of Physiologically Characterized Spherical Bushy Cell Axons from the Cochlear Nucleus of Cat: Evidence for Delay Lines to the Medial Superior Olive," *J. Comp. Neurol.*, Vol. 331, 1993, pp. 245–260.
48. G. C. Thompson and R. B. Masterton, "Brain Stem Auditory Pathways Involved in Reflexive Head Orientation," *J. Neurophysiol.*, Vol. 41, 1978, pp. 1183–1202.
49. W. M. Jenkins and M. M. Merzenich, "Role of Cat Primary Auditory Cortex for Sound-Localization Behavior," *J. Neurophysiol.*, Vol. 52, 1984, pp. 819–847.
50. J. C. Boudreau and C. Tsuchitani, "Binaural Interaction in the Cat Superior Olive S Segment," *J. Neurophysiol.*, Vol. 31, 1968, pp. 442–454.
51. D. Caird and R. Klinke, "Processing of Binaural Stimuli by Cat Superior Olivary Complex Neurons," *Exp. Brain Res.*, Vol. 52, 1983, pp. 385-399.
52. H. Held, "Die centrale Gehörleittung," *Arch. Anat. Physiol. Abt.*, Vol. 1893, pp. 201–248.
53. D. K. Morest, "The Collateral System of the Medial Nucleus of the Trapezoid Body of the Cat, Its Neuronal Architecture and Relation to the Olivo-cochlear Bundle," *Brain Res.*, Vol. 9, 1968, pp. 288–311.
54. L. P. Tolbert and D. K. Morest, "The Neuronal Architecture of the Anteroventral Cochlear Nucleus of the Cat in the Region of the Cochlear Nerve Root: Golgi and Nissl Methods," *Neuroscience*, Vol. 7, 1982, pp. 3013–3030.
55. J. J. Guinan, Jr. and R. Y. S. Li, "Signal Processing in Brainstem Auditory Neurons which Receive Giant Endings (Calyces of Held) in the Medial Nucleus of the Trapezoid Body of the Cat," *Hear. Res.*, Vol. 49, 1990, pp. 321–334.
56. S. H. Wu and J. B. Kelly, "Physiological Properties of Neurons in the Mouse Superior Olive: Membrane Characteristics and Postsynaptic Responses Studied In Vitro," *J. Neurophysiol.*, Vol. 65, 1991, pp. 230–246.
57. M. I. Banks and P. H. Smith, "Intracellular Recordings from Neurobiotin-labeled Cells in Brain Slices of the Rat Medial Nucleus of the Trapezoid Body," *J. Neurosci.*, Vol. 12, 1992, pp. 2819–2837.
58. M. J. Moore and D. M. Caspary, "Strychnine Blocks Binaural Inhibition in Lateral Superior Olivary Neurons," *J. Neurosci.*, Vol. 3, 1983, pp. 237–242.
59. N. B. Cant, "The Fine Structure of the Lateral Superior Olivary Nucleus of the Cat," *J. Comp. Neurol.*, Vol. 227, 1984, pp. 63–77.
60. K. K. Glendenning and R. B. Masterton, "Acoustic Chiasm: Efferent Projections of the Lateral Superior Olive," *J. Neurosci.*, Vol. 3, 1983, pp. 1521–1537.
61. R. L. Saint Marie, E.-M. Ostapoff, D. K. Morest, and R. J. Wenthold, "Glycine Immunoreactive Projection of the Cat Lateral Superior Olive: Possible Role in Midbrain Ear Dominance," *J. Comp. Neurol.*, Vol. 279, 1989, pp. 382–396.
62. R. Irving and J. M. Harrison, "The Superior Olivary Complex and Audition: A Comparative Study," *J. Comp. Neurol.*, Vol. 130, 1967, pp. 77–86.
63. B. Masterton, G. C. Thompson, J. K. Bechtold, and M. J. RoBards, "Neuroanatomical Basis of Binaural Phase-Difference Analysis for Sound Localization: A Comparative Study," *J. Comp. Physiol. Psychol.*, Vol. 89, 1975, pp. 379–386.
64. G. B. Henning, "Detectability of Interaural Delay in High-Frequency Complex Waveforms," *J. Acoust. Soc. Am.*, Vol. 55, 1974, pp. 84–90.
65. P. X. Joris and T. C. T. Yin, "Responses to Amplitude-Modulated Tones in the Auditory Nerve of the Cat," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 215–232.
66. P. X. Joris and T. C. T. Yin, "Envelope Coding in the Lateral Superior Olive. I. Sensitivity to Interaural Time Differences," *J. Neurophysiol.*, Vol. 73, 1995, pp. 1043–1062.
67. P. G. Finlayson and D. M. Caspary, "Low-Frequency Neurons in Lateral Superior Olive Exhibit Phase-Sensitive Binaural Inhibition," *J. Neurophysiol.*, Vol. 65, 1991, pp. 598–605.
68. L. M. Aitkin, D. J. Anderson, and J. F. Brugge, "Tonotopic Organization and Discharge Characteristics of Single Neurons in Nuclei of the Lateral Lemniscus of the Cat," *J. Neurophysiol.*, Vol. 33, 1970, pp. 421–440.
69. J. F. Brugge, D. J. Anderson, and L. M. Aitkin, "Responses of Neurons in the Dorsal Nucleus of the Lateral Lemniscus of the Cat to Binaural Tonal Stimulation," *J. Neurophysiol.*, Vol. 33, 1970, pp. 441-458.
70. M. M. Merzenich and M. D. Reid, "Representation of the Cochlea within the Inferior Colliculus of the Cat," *Brain Res.*, Vol. 77, 1974, pp. 397–415.
71. M. N. Semple and L. M. Aitkin, "Representation of Sound Frequency and Laterality by Units in Central Nucleus of Cat Inferior Colliculus," *J. Neurophysiol.*, Vol. 42, 1979, pp. 1626–1639.
72. J. E. Rose, N. B. Gross, C. D. Geisler, and J. E. Hind, "Some Neural Mechanisms in the Inferior Colliculus of the Cat which May Be Relevant to Localization of a Sound Source," *J. Neurophysiol.*, Vol. 29, 1966, pp. 288-314.
73. T. C. T. Yin and S. Kuwada, "Binaural Interaction in Low-Frequency Neurons in Inferior Colliculus of the Cat. III. Effects of Changing Frequency," *J. Neurophysiol.*, Vol. 50, 1983, pp. 1020–1042.
74. T. C. T. Yin, J. C. K. Chan, and D. R. F. Irvine, "Effects of Interaural Time Delays of Noise Stimuli on Low-Frequency Cells in the Cat's Inferor Colliculus. I. Responses to Wide-band Noise," *J. Neurophysiol.*, Vol. 55, 1986, pp. 280–300.
75. L. H. Carney and T. C. T. "Responses of Low-Frequency Cells in the Inferior Colliculus to Interaural Time Differ-

ences of Clicks: Excitatory and Inhibitory Components," *J. Neurophysiol.*, Vol. 62, 1989, pp. 144–161.

76. T. C. T. Yin, J. C. K. Chan, and L. H. Carney, "Effects of Interaural Time Delays of Noise Stimuli of Low-Frequency Cells in the Cat's Inferior Colliculus. III. Evidence for Cross-Correlation," *J. Neurophysiol.*, Vol. 1987, pp. 562–583.
77. D. R. F. Irvine and G. Gago, "Binaural Interaction in High Frequency Neurons in Inferior Colliculus of the Cat: Effects of Variations in Sound Pressure Level on Sensitivity to Interaural Intensity Differences," *J. Neurophysiol.*, Vol. 63, 1990, pp. 570–591.
78. M. N. Semple and L. M. Kitzes, "Binaural Processing of Sound Pressure Level in the Inferior Colliculus," *J. Neurophysiol.*, Vol. 57, 1987, pp. 1130–1147.
79. T. C. T. Yin, S. Kuwada, and Y. Sujaku, "Interaural Time Sensitivity of High-Frequency Neurons in the Inferior Colliculus," *J. Acoust. Soc. Am.*, Vol. 76, 1984, pp. 1401–1410.
80. A. D. Musicant, J. C. K. Chan, and J. E. Hind, "Direction-Dependent Spectral Properties of Cat External Ear: New Data and Cross-Species Comparisons," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 757–781.
81. M. Konishi, T. T. Takahashi, H. Wagner, W. E. Sullivan, and C. E. Carr, "Neurophysiological and Anatomical Substrates of Sound Localization in the Owl," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function: The Neurobiological Bases of Hearing*, Wiley, New York, 1988, pp. 721–745.
82. D. R. Moore, M. N. Semple, P. D. Addison, and L. M. Aitkin, "Properties of Spatial Receptive Fields in the Central Nucleus of the Cat Inferior Colliculus. I. Responses to Tones of Low Intensity," *Hear. Res.*, Vol. 13, 1984, pp. 159–174.
83. L. M. Aitkin, J. D. Pettigrew, M. B. Calford, S. C. Phillips, and L. Z. Wise, "Representation of Stimulus Azimuth by Low-Frequency Neurons in Inferior Colliculus of the Cat," *J. Neurophysiol.*, Vol. 53, 1985, pp. 43–59.
84. M. N. Semple, L. M. Aitkin, J. D. Pettigrew, M. B. Calford, and D. P. Phillips, "Spatial Receptive Fields in the Cat Inferior Colliculus," *Hear. Res.*, Vol. 10, 1983, pp. 203–215.
85. D. L. Sparks, "Translation of Sensory Signals into Commands for Control of Saccadic Eye Movements: Role of Primate Superior Colliculus," *Physiol. Rev.*, Vol. 66, 1986, pp. 118–171.
86. L. M. Aitkin and W. R. Webster, "Medial Geniculate Body of the Cat: Organization and Responses to Tonal Stimuli of Neurons in Ventral Division," *J. Neurophysiol.*, Vol. 35, 1972, pp. 365–380.
87. M. B. Calford and W. R. Webster, "The Auditory Representation within the Principal Division of the Cat Medial Geniculate Body: An Electrophysiological Study," *J. Neurophysiol.*, Vol. 45, 1981, pp. 1013–1028.
88. T. J. Imig and A. Morel, "Tonotopic Organization in Ventral Nucleus of Medial Geniculate Body in the Cat," *J. Neurophysiol.*, Vol. 53, 1985, pp. 309–340.
89. E. M. Rouiller and F. de Ribaupierre, "Origin of Afferents to Physiologically Defined Regions of the Medial Geniculate Body of the Cat," *Hear. Res.*, Vol. 19, 1985, pp. 97–114.
90. J. E. Rose, "The Cellular Structure of the Auditory Region of the Cat," *J. Comp. Neurol.*, Vol. 91, 1949, pp. 409–439.
91. J. E. Rose and C. N. Woolsey, "The Relations of Thalamic Connection, Cellular Structure and Evokable Electrical Activity in the Auditory Regions of the Cat," *J. Comp. Neurol.*, Vol. 91, 1949, pp. 441–466.
92. J. F. Brugge and R. A. Reale, "Auditory Cortex," in A. Peters and E. G. Jones (Eds.), *Cerebral Cortex, Vol. 4, Association and Auditory Cortices*, Plenum, New York, 1985, pp. 229–271.
93. R. A. Reale and T. J. Imig, "Tonotopic Organization in the Auditory Cortex of the Cat," *J. Comp. Neurol.*, Vol. 192, 1980, pp. 265–292.
94. T. J. Imig and H. O. Adrian, "Binaural Columns in the Primary Field (AI) of Cat Auditory Cortex," *Brain Res.*, Vol. 138, 1977, pp. 241–257.
95. T. J. Imig, W. A. Irons, and F. R. Samson, "Single-Unit Selectivity to Azimuthal Direction and Sound Pressure Level of Noise Bursts in Cat High-Frequency Primary Auditory Cortex," *J. Neurophysiol.*, Vol. 63, 1990, pp. 1448–1466.
96. R. Rajan, L. M. Aitkin, and D. R. F. Irvine, "Azimuthal Sensitivity of Neurons in Primary Auditory Cortex of Cats. II. Organization along Frequency-Band Strips," *J. Neurophysiol.*, Vol. 64, 1990, pp. 888–902.
97. J. F. Brugge, N. A. Dubrovsky, L. M. Aitkin, and D. J. Anderson, "Sensitivity of Single Neurons in Auditory Cortex of Cat to Binaural Tonal Stimulation: Effects of Varying Interaural Time and Intensity," *J. Neurophysiol.*, Vol. 32, 1969, pp. 1005–1024.
98. D. P. Phillips and D. R. F. Irvine, "Some Features of Binaural Input to Single Neurons in Physiologically-Defined Area AI of Cat Cerebral Cortex," *J. Neurophysiol.*, Vol. 49, 1983, pp. 383–395.
99. R. A. Reale and J. F. Brugge, "Auditory Cortical Neurons Are Sensitive to Static and Continuously Changing Interaural Phase Cues," *J. Neurophysiol.*, Vol. 64, 1990, pp. 1247–1260.
100. R. Rajan, L. M. Aitkin, D. R. F. Irvine, and J. McKay, "Azimuthal Sensitivity of Neurons in Primary Auditory Cortext of Cats. I. Types of Sensitivity and the Effects of Variations in Stimulus Parameters," *J. Neurophysiol.*, Vol. 64, 1990, pp. 872–887.
101. E. Borg, "On the Neuronal Organization of the Acoustic Middle Ear Reflex. A Physiological and Anatomical Study," *Brain Res.*, Vol. 49, 1973, pp. 101–123.
102. J. J. Guinan, Jr., M. P. Joseph, and B. E. Norris, "Brain-

stem Facial-Motor Patterns from Two Distinct Groups of Stapedius Motoneurons in the Cat," *J. Comp. Neurol.*, Vol. 287, 1989, pp. 134–144.

103. S. R. Vacher, J. J. Guinan, Jr., and J. B. Kobler, "Intracellularly Labeled Stapedius Motoneuron Cell Bodies in the Cat Are Spatially Organized According to Their Physiological Responses," *J. Comp. Neurol.*, Vol. 289, 1989, pp. 401–415.
104. W. B. Warr and J. J. Guinan, Jr., "Efferent Innervation of the Organ of Corti: Two Separate Systems," *Brain Res.*, Vol. 173, 1979, pp. 152–155.
105. A. R. Cody and B. M. Johnstone, "Temporary Threshold Shift Modified by Binaural Acoustic Stimulation," *Hear. Res.*, Vol. 6, 1982, pp. 199–205.
106. R. L. Winslow and M. B. Sachs, "Effects of Electrical Stimulation of the Crossed Olivocochlear Bundle on Auditory Nerve Response to Tones in Noise," *J. Neurophysiol.*, Vol. 57, 1987, pp. 1002–1021.

112

BIOCHEMISTRY AND PHARMACOLOGY OF THE AUDITORY SYSTEM

WILLIAM F. SEWELL

1 INTRODUCTION

The field of biochemistry and pharmacology of the inner ear is poised for great change and advance with the advent of widely available molecular biological approaches. The following is an extremely brief synopsis of the pharmacology and biochemistry of the auditory system as it stands in 1991. It is intended to provide an overview for the nonspecialist and to provide for specialist and nonspecialist alike as much data in tabular form as is feasible. More detailed presentation of this field can be found in a number of excellent reviews cited throughout the text. In general, appropriate reviews are cited rather than original references, which may help guide the reader to the available literature. The biochemical data are organized principally around the presentation of measurements of metabolites and enzymes within the inner ear. The pharmacological presentation takes a practical viewpoint and describes drugs that might be useful in attenuating various physiological processes in the inner ear. Scientists studying the auditory system who may want to selectively eliminate functional elements in the cochlea might find this useful. Neurotransmitters are described in simple terms as a broad guide for the nonexpert. Coverage of the central auditory nervous system is limited to a description of neurotransmitters in the cochlear nucleus.

2 BIOCHEMISTRY OF THE COCHLEA

2.1 Chemical Composition of the Cochlear Fluids

The two specialized types of extracellular fluid in the inner ear are perilymph and endolymph. A description of the chemical composition of the cochlear fluids is provided in Table 1. Perilymph fills the scala vestibuli and scala tympani and bathes most of the structures of the organ of Corti, including the dendritic endings of the auditory nerve fibers, the basolateral surfaces of the hair cells, and the spiral ligament. The perilymph of scala tympani is in diffusional continuity with the cerebrospinal fluid (CSF) via the cochlear aqueduct and is generally similar in ionic composition to CSF. Thalmann[2] suggests that many of the published measurements of perilymph chemical content probably reflect contamination of the perilymph with CSF. The amino acids glycine and alanine are particularly high in perilymph compared to CSF. The perilymph of scala vestibuli is similar to that of scala tympani, but there are differences. In particular, perilymph in scala vestibuli has higher potassium and protein concentrations.

Endolymph fills the scala media and bathes the apical surfaces of the hair cells. Endolymph is similar in cationic composition to intracellular fluid, having a high K^+ concentration and a low Na^+ concentration. The scala media has an electrical potential of about +100 mV relative to perilymph and blood. The cells that line the scala media are connected with tight junctions that

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

TABLE 1 Chemical Composition of Cochlear Fluids

	CSF	Scala Tympani	Scala Vestibuli	Scala Media
Na^+ (mM)	149	147	141	1.4
K^+ (mM)	2.9	3.4	6.7	161
Cl^- (mM)	134	125	129	131
Ca^{2+} (mM)		.64	.68	0.045
Mg^{2+} (mM)		1		.2
pH	7.3	7.3	7.3	7.4
Osmolality (mOsm/kg H_2O)		293	294	304
Glucose	80	80	80	11
Protein (mg/100 ml)	32	180	237	30
Alanine	36	73–347	367	40
Arginine	26	20–46	35	4
Asparagine	2.4	2–4	3.7	.6
Aspartate	3	2.5–4	3.4	15
Glutamate	3	5–9	8	53
Glutamine	420	428–705	604	54
Glycine	12	230–250	247	71
Histidine	19	27–39	27.6	1
Isoleucine	9	19	15.6	2
Leucine	15	20–43	30	2.5
Lysine	43	50–69	64.6	5
Methionine	11	14–50	33	2.7
Ornithine	10	13–31	28.4	5.5
Phenylalanine	10	9–32	26.2	2
Serine	65	76–154	123	19
Taurine	6	7.5		
Threonine	25	60–65	63	15
Tryptophan	5	6–10	12	2
Tyrosine	18	29–43	40	2.6
Valine	22	39–60	53.8	3.2

Note: All concentrations expressed in micromolar unless otherwise indicated. Measurements for amino acids and protein are from Refs. 1 and 2. Measurements of other constituents are from reviews in Refs. 3–5.

help contain the unique electrochemical environment of the scala media. There are other chemical differences between endolymph and perilymph. Protein concentration is higher in perilymph. Acidic amino acids are about 5 times higher in endolymph than in perilymph. Neutral and basic amino acid concentrations in endolymph are $\frac{1}{3}$ to $\frac{1}{15}$ those of perilymph.

2.2 Cochlear Metabolism

The stria vascularis is a richly vascularized structure on the lateral wall of the cochlea important in creating the unique electrochemical environment of the scala media. The marginal cells, which line the border between the stria vascularis and the endolymph, are thought to generate the endocochlear potential. They have positive intracellular potentials that are equal to, or slightly above, that of the scala media.[6–8] The basal poles of the marginal cells form intimate contacts with the blood vessels of the stria vascularis and contain large numbers of mitochondria.

A large number of enzyme activities and metabolite concentrations have been quantified in rapidly frozen and freeze-dried microdissected cochlear structures. These data are summarized in Tables 2 and 3. Both the stria vascularis and the organ of Corti have high rates of metabolism. The stria vascularis relies heavily on the aerobic metabolism of glucose and has been compared, metabolically, to slow oxidative muscle.[12] This energy is used to generate the endocochlear potential and to contribute to the ionic environment (high potassium/low sodium) of the scala media. The organ of Corti has a very high level of total energy reserves with the outer hair cells containing high levels of glycogen and enzymes of glycogen metabolism. Glycogen levels are much higher in the apex than in the base of the cochlea.[15] Metaboli-

TABLE 2 Enzyme Activities in Cochlea

	Organ of Corti	Inner Hair Cell Region	Outer Hair Cell Region	Stria Vascularis	Reissner's Membrane	Brain
Acetylcholinesterase		4.4	2.8	0.026	0.02	2.3
Adenylate cyclase	12–16			19–22	<0.5	
Aspartate aminotransferase		13	11.2	20.3	3.5	46
Carbonic anhydrase	47			286		
Choline acetyltransferase		0.056	0.019	0.00006	0.00014	0.015
Glucose-6-*P*-dehydrogenase	1.2			1.1	0.32	0.3
Glutamate decarboxylase	0.17					
Glycerol-1-*P*-dehydrogenase	0.1			0.11	0.04	
Hexokinase	0.47			0.27	0.093	1.6
Lactate dehydrogenase	18–22	20	33	15–18	0.6	19
Malate dehydrogenase	4.9–5.8	7.4	6.8	12.5–14.2	1.94	30.2
Mg–ATPase	1–2.8			2–6	1	
Na,K–ATPase	0.6			7.3	0.4	
Phosphoglucomutase	13.8			9.0	2.6	
6-*P*-gluconate dehydrogenase	0.38			0.14	0.05	0.16–0.3
Total phosphorylase	0.74			0.16	0.1	
Phosphorylase a	0.17			0.05	0.01	
Phosphofructokinase	2.8			2.3	0.13	2.6
Pyruvate kinase	4.9			4.8	2.1	
UDPG–pyrophosphorylase	4.9			2.5	1.4	

Note: Enzyme activities are expressed as mol substrate converted/kg dry tissue/h, except for carbonic anydrase and glutamate decarboxylase in which activity is expressed as units per mg protein. 1 unit of carbonic anydrase is the amount of carbonic anhydrase that increased the reaction rate 10% above that of the uncatalyzed reaction. 1 unit of glutamate decarboxylase is 1 nanomole GABA formed/min. Data are from Refs. 9–14.

TABLE 3 Metabolite Concentrations in Cochlea

		Organ of Corti	Inner Hair Cell Region	Outer Hair Cell Region	Stria Vascularis	Reissner's Membrane	Brain
Glycogen	mmoles/kg dry wt	100–300	100	500–600	67		
ATP	mmoles/kg dry wt	15.9			10.9	4.6	
Creatine	mmoles/kg	9.2			6.4	2.5	
Phosphate	dry wt	37					
Glucose	mmoles/kg dry wt				10.5		
c-AMP	μmoles/kg dry wt		4.2	4.3			
c-GMP	nmoles/kg dry wt		180–300	40–230			
Glutamate	mmoles/kg dry wt	60–81	59	74	25	9.6	36–54
Aspartate[a]	mmoles/kg dry wt	52–109	19.1	22.1	3.6	2.2	10–15
Glycine	mmoles/kg dry wt		35	37	33	39	4.4–6.8
GABA	mmoles/kg dry wt		2.2	1.7	1.9	2.1	8.4–9.2

Note: Values expressed per kg dry wt. may be divided by 5 to give a rough estimate of the concentration in tissue. Data are from Refs. 10, 12, and 38.

[a] Aspartate values in organ of Corti are from chinchilla while those of selected cochlear regions are from guinea pig.

cally, the hair cells have been compared to "fast" glycolytic skeletal muscle.[12]

The electrochemical environment of the scala media allows the hair cells to transduce acoustic stimulation by modulating a steady-state current flowing from scala media through the hair cells. Most of this current is carried by potassium, which exits the basolateral surface of the hair cell down an electrochemical gradient. Thus energy expenditure in the stria vascularis relieves the hair cells of expending a great deal of energy in ionic balance during transduction. Thalmann et al.[16] have shown that availability of energy substrate is not a factor in organ of Corti pathophysiology.

2.3 Biochemistry of Tectorial Membrane

The tectorial membrane is an extracellular gel consisting of fibrils embedded in an amorphous substance. One fourth to one half of the tectorial membrane protein appears to be collagen (possibly types II, IX, and XI) and nearly half of the total protein is accounted for by at least three noncollagenous glycosylated polypeptides.[17–19]

2.4 Proteins in Hair Cells

A number of structural elements in hair cells have been described (reviewed in Ref. 20). The cilia are composed of actin filaments organized in a paracrystalline array with crossbridges formed by the actin-binding protein fimbrin. The cuticular plate contains densely packed myosin and actin. Stereocilia contain at least two calcium-binding proteins, calbindin and calmodulin. Several other as-yet-unidentified proteins are present in stereocilia and the organ of Corti in relatively high amounts.[21,22]

3 PHARMACOLOGY OF THE COCHLEA

Drugs that affect the auditory system are useful as probes of auditory function and may alter certain processes with some selectivity. However, it must be remembered that the selectivity of a drug decreases with increasing concentration. This section is generally organized by the functional processes in the cochlea that can be affected with some selectivity by drugs.

3.1 Blood–Perilymph Barrier

Access of drugs to the cochlea is restricted by the presence of the blood–cochlea barrier,[23] similar in many ways to the blood–brain barrier. Most of the structures of the cochlea are bathed in perilymph and in endolymph. Drugs that are highly charged do not readily enter the fluid compartments of the cochlea from the systemic circulation and must be instilled into the cochlea to act. Access of positively charged drugs to the endolymph is further restricted because of the relatively large (+100 mV) electrical potential in the endolymph. Drugs that can enter the perilymph and are highly protein bound can concentrate in the inner ear due to the relatively high concentration of protein in the perilymph. One cochlear structure that is readily accessible from the systemic circulation is the stria vascularis. Some drugs placed in the middle ear can diffuse across the round window membrane to gain access to the perilymph; this route of administration is not yet well characterized.

3.2 Drugs That Affect Endolymphatic Potential

Loop Diuretics The loop diuretics, also called high-ceiling diuretics, are a class of drugs that produce diuresis by inhibiting electrolyte reabsorption in the loop of Henle. These agents include furosemide, piretanide, and bumetanide. They suppress cochlear responses by reversibly decreasing the endocochlear potential. All of these drugs act rapidly after intravenous administration and the effects correlate with plasma concentrations of the drug. They apparently act by blocking ion transport in the marginal cells of the stria vascularis.[24]

Ouabain Ouabain is a glycoside that inhibits Na-K-ATPase, an enzyme that regulates the sodium and potassium concentrations in cells. While all cells have some Na-K-ATPase, the stria vascularis contains large amounts. Perfusion of the perilymph with ouabain reduces the endolymphatic potential and cochlear microphonic potential,[25] but the effects of ouabain do not actually appear to be selective for any particular cell population in the cochlea and may reflect a general compromise of cellular ability to regulate ionic balance.

Metabolic Inhibitors Another class of drugs that alter the endocochlear potential are metabolic inhibitors, such as cyanide and 2,4-dinitrophenol.[16] These drugs will affect all processes that rely on aerobic metabolism and are not necessarily selective for actions on endolymph production. However, because the stria vascularis has such a high metabolic rate and is in diffusional contact with the systemic circulation, effects on endolymphatic potential are often more dramatic than those on other cochlear processes.

3.3 Drugs That Affect Hair Cells

Transduction Channel Blockers Transduction in the hair cells requires movement of ions through channels at the tips of the stereocilia. The channel is a relatively nonselective cation channel and can be blocked by perfusion of the apical surface of the hair cell with tetramethylammonium or by lowering endolymph calcium concentration to less than 10 μM.[26] The aminoglycoside antibiotics are known to block transduction,[27] presumably by virtue of their properties as large cations.

Glycolysis Inhibitors Iodoacetate, a powerful inhibitor of glycolysis, can reduce the amplitude of the cochlear microphonic potential when perfused into the perilymphatic space.[16] Given the very high levels of glycogen in the outer hair cells, this action may reflect a disruption of fast metabolic energy use in the outer hair cell.

Agents That Destroy Hair Cells The aminoglycoside antibiotics can produce destruction of the hair cells. These antibiotics are a group of polycationic glycosaminoglycan-based compounds used clinically in the treatment of infection by gram-negative bacteria. Some common ototoxic aminoglycoside antibiotics are neomycin, dihydrostreptomycin, streptomycin, kanamycin, gentamicin, and amikacin. The ototoxic effects are seen at doses that produce toxic effects elsewhere in the body. Hair cells are most susceptible to the destruction by these antibiotics, although other effects, such as destruction of the stria vascularis and of supporting cells in the cochlea, are also observed.[28] The toxic effects of the aminoglycosides are potentiated by concomitant administration of ethacrinic acid, a diuretic that can reduce the endocochlear potential (see Ref. 29 for a review). The mechanism by which aminoglycosides act to destroy hair cells is not clear, although many plausible hypotheses have been put forth (see Ref. 30 for a review).

3.4 Drugs That Affect Afferent Synaptic Transmission

Calcium Channel Blockers The release of neurotransmitter from the hair cells requires an influx of calcium via voltage-sensitive calcium channels. Thus synaptic transmission can be blocked by preventing the entry of calcium through these channels. Agents used for this purpose have included cobalt, active at concentations of around 1 mM, and magnesium, active at around 20 mM.[31,32]

Excitatory Amino Acid Antagonists Neurotransmitter released by the hair cell acts on a postsynaptic receptor thought to be an excitatory amino acid receptor. Antagonists for those receptors can block afferent transmission. The most potent blockers are those that block non-NMDA receptors and include the quinoxalinediones, DNQX and CNQX.[33]

3.5 Drugs That Block Action Potential Propagation

Tetrodotoxin is a potent blocker of the regenerative sodium channels in neurons. Tetrodotoxin administration can block the generation of action potentials in all nerve fibers in the cochlea, including afferent and efferent nerve fibers. Mammalian hair cells do not possess regenerative sodium channels and are not affected.

3.6 Drugs That Block Middle Ear Muscles

The motoneurons to the middle ear muscles use acetylcholine as a neurotransmitter, as do all vertebrate motoneurons. The cholinergic receptors appear to be muscle nicotinic receptors and can be blocked by anticholinergic paralytic agents, although they may be atypical in their responses to these drugs.[34]

3.7 Miscellaneous Ototoxic Agents

Other ototoxic agents include aspirin, quinine, and cisplatinum. The mechanisms of ototoxicity of these agents are not well understood.

4 NEUROTRANSMITTERS IN THE COCHLEA

4.1 Afferent Transmission

The hair cell releases a neurotransmitter to excite afferent nerve fibers. The neurotransmitter has not been identified, although most common candidates are ruled out on the basis of pharmacological, biochemical, or immunohistochemical experiments.[35,36] Glutamate is often proposed to be the hair cell transmitter, but because of its ubiquity in presence and action, glutamate has been difficult to prove to be the hair cell transmitter. While glutamate concentration in the organ of Corti is relatively high, 12–15 mM, concentrations in inner hair cells are actually lower than those in supporting cells or outer hair cells.[37,38] Glutamate applied to the perilymph can excite afferent fibers although the effects are seen only at high (mM) concentrations. While the question of whether glutamate is the afferent neurotransmitter is unresolved, it is becoming evident that the effects of the hair cell transmitter are probably mediated via excitatory amino acid receptors on afferent fibers.

4.2 Efferent Transmission

The cochlea is innervated by two groups of efferent nerve fibers: Large myelinated efferent fibers originate near the medial nucleus of the trapezoid body and synapse predominantly on outer hair cells; small unmyelinated efferent fibers originate in the hilus of the lateral superior olivary complex and terminate predominantly on the radial afferent fibers that innervate the inner hair cells. These two efferent groups are respectively termed the medial and lateral efferents. The major neurotransmitter used by the medial efferent fibers is acetylcholine.[39] The lateral efferents contain, in addition to acetylcholine, an assortment of neuropeptides, including enkephalins, dynorphins, and CGRP.[40] The actions of these peptides in the cochlea are, for the most part, unknown, although opioid receptor activation suppresses adenylate cyclase activity in the cochlea.[41]

The major effects of efferent stimulation can be blocked by many cholinergic antagonists provided the antagonists are applied to the perilymph. The effects of efferent activity are also blocked by strychnine, an agent often thought of as a glycinergic blocker, although it is known to block cholinergic transmission at other synapses.

5 NEUROTRANSMITTERS IN THE COCHLEAR NUCLEUS

5.1 Gamma-aminobutyric acid

Gamma-aminobutyric acid (GABA) concentration in the cochlear nucleus is about one third that of the average brain concentration. Most of the GABA and its associated metabolic enzymes are concentrated in the dorsal cochlear nucleus (DCN).[42] Two thirds of the GABA in the cochlear nucleus is associated with interneurons.[43] Putative GABAergic terminals tend to contain pleomorphic synaptic vesicles and symmetrical synaptic junctions.[44] Suspected GABAergic systems within the cochlear nucleus include: stellate cells of DCN to granule cells of DCN, deep DCN cells to the anteroventral cochlear nucleus (AVCN), periolivary cells of superior olivary complex (SOC) to ventral cochlear nucleus (VCN), bipolar cells of DCN to contralateral DCN, and cartwheel cells of DCN to pyramidal cells of the DCN. Cochlear nucleus neurons with substantial GABAergic input include octopus cells, spherical bushy cells of AVCN, pyramidal cells of the DCN, and granule cells of DCN.[44–46]

5.2 Glycine

Glycine concentration in the cochlear nucleus is twice that of average brain concentration.[42] Much of the glycine in the cochlear nucleus is associated with interneurons, input from the contralateral cochlear nucleus, and input from the ventral nucleus of the trapezoid body.[47] Because glycine has many prominent metabolic roles, including production of proteins, purines, and other amino acids, it is difficult to separate neurotransmitter from metabolic pools of glycine. Strychnine binding, taken to be an indicant of the presence of glycinergic synapses, is highest in DCN, suggesting that glycinergic transmission may be most prevalent in the DCN.[48] However, most of the principal cells of the VCN apparently also receive some glycinergic input.[47,49] Glycinergic terminals are thought to contain flattened vesicles.[47] The cartwheel cells of DCN are thought to inhibit pyramidal of the DCN via glycinergic synapses.[49] That this circuit is considered both GABAergic (see above) and glycinergic pointedly illustrates the problem of inferring neurotransmitter function from immunohistochemical analysis.

5.3 Acetylcholine

Two major sources of cholinergic input in the cochlear nucleus are OCB collaterals to multipolar cells near the granule cell layer of the VCN and ventral centrifugal projections from in or near the trapezoid body. In some species there are choline acetyltransferase positive cells in the cochlear nucleus indicating the presence of some possibly cholinergic neurons.[50]

5.4 Noradrenaline

Most cell types of the AVCN and PVCN receive some adrenergic input via bilateral projections from the region of the locus coeruleus.[51]

5.5 Glutamate or Related Substances

Glutamate concentrations are uniform within the cochlear nucleus and are about the same as in whole brain. Aspartate, on the other hand, is lowest in the granule cell layer and highest in areas innervated by the auditory nerve.[42] As with glycine, glutamate and aspartate have a number of prominent metabolic roles, including protein formation and ammonium transfer. It is not presently possible to separate metabolic pools of these amino acids from neurotransmitter pools. Many, but not all, neurons in the cochlear nucleus are excited by glutamate and its analogs.[49,52,53] Glutamate or a related substance has been proposed as the major fast excitatory transmitter in the central nervous system. Within the cochlear nucleus, auditory nerve fibers, and granule cells are proposed to be glutamatergic.[54]

REFERENCES

1. R. Thalmann, T. H. Comegys, and I. Thalmann, "Amino Acid Profiles in Inner Ear Fluids and Cerebrospinal Fluid," *Laryngoscope*, Vol. 92, 1982, pp. 321–328.

2. R. Thalmann, "Gradients of Organic Substances between Fluid Compartments of the Cochlea," *Acta Otolaryngo.*, Vol. 99, 1985, pp. 469–477.
3. A. N. Salt and T. Konishi, "The Cochlear Fluids: Perilymph and Endolymph," in R. A. Altschuler, R. P. Bobbin, and D. W. Hoffman (Eds.), *Neurobiology of Hearing: The Cochlea*, Raven Press, New York, 1986.
4. A. N. Salt and R. Thalmann, "Cochlear Fluid Dynamics," in A. F. Jahn and J. Santos-Sacchi (Eds.), *Physiology of the Ear*, Raven Press, New York, 1988.
5. O. Sterkers, E. Ferrary, and C. Amiel, "Production of Inner Ear Fluids," *Physiol. Rev.*, Vol. 98, 1988, pp. 1083–1128.
6. I. Melichar and J. Syka, "DC Potentials in Different Cells of the Stria Vascularis Measured In Vitro," *Hear. Res.*, Vol. 25, 1987, pp. 23–33.
7. F. F. Offner, P. Dallos, and M. A. Cheatham, "Positive Endocochlear Potential: Mechanism of Production by Marginal Cells of Stria Vascularis," *Hear. Res.*, Vol. 29, 1987, pp. 117–124.
8. J. Prazma, "Electroanatomy of the Lateral Wall of the Cochlea," *Arch. Oto-Rhino-Laryngol.*, Vol. 209, 1975, pp. 1–33.
9. I. Thalmann, R. Matchinsky, and R. Thalmann, "Quantitative Study of Selected Enzymes Involved in Energy Metabolism of the Cochlear Duct," *Ann. Otol. Rhinol. Laryngol*, Vol. 79, 1970, pp. 1–8.
10. R. Thalmann, "Metabolic Features of Auditory and Vestibular Systems," *Laryngoscope*, Vol. 81, 1971, pp. 1245–1260.
11. S. Paloheimo and R. Thalmann, "Influence of Loop Diuretics upon Na^+K^+ATPase and Adenylate Cyclase of the Stria Vascularis," *Arch. Oto-Rhino-Laryngol*, Vol. 217, 1977, pp. 347–359.
12. D. A. Godfrey, G. J. Wiet, and C. D. Ross, "Quantitative Histochemistry of the Cochlea," in R. A. Altschuler, R. P. Hoffman, and R. P. Bobbin (Eds.), *Neurobiology of Hearing*, Raven Press, New York, 1986.
13. J. Fex and R. J. Wenthold, "Choline Acetyltransferase, Glutamate Decarboxylase, and Tyrosine Hydroxylase in the Cochlea and Cochlear Nucleus of the Guinea Pig," *Brain Res.*, Vol. 421, 1976, pp. 336–342.
14. D. G. Drescher and T. P. Kerr, "Na^+K^+ Activated Adenosine Triphophatase and Carbonic Anhydrase: Inner Ear Enzymes of Ion Transport," in D. G. Drescher (Ed.), *Auditory Biochemistry*, Thomas, Springfield, IL, 1985.
15. R. Thalmann, T. Miyoshi, and I. Thalmann, "The Influence of Ischemia upon the Energy Reserves of Inner Ear Tissue," *Laryngoscope*, Vol. 87, 1972, pp. 2249–2272.
16. R. Thalmann, I. Thalmann, I. Ise, and S. Paloheimo, "Noxious Effects upon Cochlear Metabolism," *Laryngoscope*, Vol. 87, 1977, 699–721.
17. G. P. Richardson, I. J. Russell, V. C. Duance, and A. J. Bailey, "Polypeptide Composition of the Mammalian Tectorial Membrane," *Hear Res.*, Vol. 25, 1987, pp. 45–60.
18. K. P. Steele, "Composition and Properties of the Mammalian Tectorial Membrane," in D. Drescher (Ed.), *Auditory Biochemistry*, Thomas, Springfield, IL, 1985.
19. I. Thalmann, G. Thallinger, T. H. Comegys, and R. Thalmann, "Collagen—the Predominant Protein of the Tectorial Membrane," *Otorhinolaryngology*, Vol. 48, 1986, pp. 107–115.
20. J. T. Corwin and M. E. Warchol, "Auditory Hair Cells: Structure, Function, Development, and Regeneration," *Ann. Rev. Neurosci.*, Vol. 14, 1991, pp. 301–333.
21. G. M. G. Shepherd, B. A. Barres, and D. P. Corey, " 'Bundle Blot' Purification and Initial Protein Characterization of Hair Cell Stereocilia," *Proc. Natl. Acad. Sci. USA*, Vol. 86, 1989, pp. 4973–4977.
22. I. Thalmann, K. Takahashi, J. Varghese, T. H. Comegys, and R. Thalmann, "Biochemical Features of Major Organ of Corti Proteins (OCP-I and OCP-II) Including Partial Amino Acid Sequence," *Laryngoscope*, Vol. 100, 1990, pp. 99–103.
23. R. Klinke and N. Galley, "Efferent Innervation of Vestibular and Auditory Receptors," *Physiol. Rev.*, Vol. 54, 1974, pp. 316–357.
24. L. Rybak, "Ototoxic Mechanisms," in R. A. Altschuler, D. W. Hoffman, and R. P. Bobbin (Eds.), *Neurobiology of Hearing: The Cochlea*, Raven Press, New York, 1986.
25. W. Kuijpers and S. L. Bonting, "The Cochlear Potentials. I: The Effect of Ouabain on the Cochlear Potentials of the Guinea Pig," *Pflügers Archives*, Vol. 320, 1970, pp. 348–358.
26. D. P. Corey and A. J. Hudspeth, "Ionic Basis of the Receptor Potential in a Vertebrate Hair Cell," *Nature*, Vol. 281, 1979, pp. 675–677.
27. T. Konishi, "Effects of Local Application of Ototoxic Antibiotics on Cochlear Potentials in Guinea Pigs," *Acta Otolaryngol.*, Vol. 88, 1979, pp. 41–46.
28. J. Hawkins, "Drug Ototoxicity," in W. D. Keidel and W. D. Neff (Eds.), *Handbook of Sensory Physiology: Clinical and Special Topics*, Springer Verlag, Berlin, 1976.
29. J. Prazma, "Ototoxicity of Aminoglycoside Antibiotics," in R. D. Brown and E. A. Daigneault (Eds.), *Pharmacology of Hearing: Experimental and Clinical Bases*, Wiley, New York, 1987.
30. J. Prazma, "Ototoxicity of Diuretics," in R. D. Brown and E. A. Daigneault (Eds.), *Pharmacology of Hearing: Experimental and Clinical Bases*, Wiley, New York, 1981.
31. W. F. Sewell, "Synaptic Potentials in Afferent Fibers Innervating Hair Cells of the Lateral Line Organ of Xenopus Laevis," *Hear. Res.*, Vol. 44, 1990, pp. 71–82.
32. J. H. Siegel and E. M. Relkin, "Antagonistic Effects of Perilymphatic Calcium and Magnesium on the Activity of Single Cochlear Afferent Neurons," *Hear. Res.*, Vol. 28, 1987, pp. 131–147.
33. T. Littman, R. P. Bobbin, M. Fallon, and J.-L. Puel, "The Quinoxalinediones DNQX and CNQX and Two Related Congeners Suppress Hair Cell-to-Auditory Nerve Transmission," *Hear. Res.*, Vol. 40, 1989, pp. 45–54.
34. L. Gisselsson, B. Lofstrom, and O. Metz, "The Effect of Succinylcholine on the Middle Ear Muscles and on the Transmission of Sound through the Middle Ear," *Acta Otolaryngolog.*, 1956, pp. 233–244.

35. P. S. Guth, M. Tachibana, and W. F. Sewell, "The Pharmacology of the Cochlear Afferents and Cochlear Nucleus," in R. S. Brown and E. A. Daigneault (Eds.), *Pharmacology of Hearing*, Wiley, New York, 1981, pp. 99–136.
36. R. Klinke, "Neurotransmission in the Inner Ear," *Hear. Res.*, Vol. 22, 1986, pp. 235–244.
37. D. A. Godfrey, J. A. Carter, S. J. Berger and F. M. Matschinsky, "Levels of Putative Transmitter Amino Acids in the Guinea Pig Cochlea," *J. Histochem. Cytochem.*, Vol. 24, No. 2, 1976, pp. 486–472.
38. I. Thalmann, R. Thalmann, D. C. Marcus, and T. H. Comegys, "Levels of Putative Neurotransmitters, Glutamate and Asparatate, in Cochlear Fluids and Cell Layers of Organ of Corti," *J. Acoust. Soc. Am.*, Vol. 66, 1979, p. S47.
39. P. S. Guth, C. H. Norris, and R. P. Bobbin, "The Pharmacology of Transmission in the Peripheral Auditory System," *Pharmacol. Rev.*, Vol. 28, 1976, pp. 95–125.
40. R. A. Altschuler, D. W. Hoffman, and R. Wenthold, "Neurotransmitters of the Cochlea and Cochlear Nucleus: Immunocytochemical Evidence," *Am. J. Otol.*, Vol. 7, 1986, pp. 100–106.
41. M. Eybalin, R. Pujol, and J. Bockaert, "Opioid Receptors Inhibit the Adenylate Cyclase in Guinea Pig Cochleas," *Brain Res.*, Vol. 421, 1987, pp. 336–342.
42. D. A. Godfrey, J. A. Carter, S. J. Berger, O. H. Lowry, and F. M. Matchinsky, "Quantitative Histochemical Mapping of Candidate Transmitter Amino Acids in Cat Cochlear Nucleus," *J. Histochem. Cytochem.*, Vol. 25, 1977, pp. 417–431.
43. S. Potashner, D. K. Morest, D. L. Oliver, and D. R. Jones, "Identification of Glutamatergic and Aspartatergic Pathways in the Auditory System," in D. G. Drescher (Ed.), *Auditory Biochemistry*, Thomas, Springfield, IL, 1985.
44. R. L. Saint Marie, E. M. Ostapoff, D. K. Morest, and R. J. Wenthold, "Glycine-Immunoreactive Projection of the Cat Lateral Superior Olive: Possible Role in Midbrain Ear Dominance," *J. Comp. Neurol.*, Vol. 279, 1989, pp. 382–396.
45. J. C. Adams and E. Mugnaini, "Patterns of Glutamate Decarboxylase Immunostaining in the Feline Cochlear Nuclear Complex Studied with Silver Enhancement and Electron Microscopy," *J. Comp. Neurol.*, Vol. 262, 1987, pp. 375–401.
46. K. K. Osen, O. P. Ottersen, and J. Storm-Mathisen, "Colocalization of Glycine-like and GABA-like Immunoreactivities: A Semiquantitative Study of Individual Neurons in the Dorsal Cochlear Nucleus of Cat," in O. P. Ottersen and J. Storm-Mathisen (Eds.), *Glycine Neurotransmission*, Wiley, Chichester, England, 1990.
47. R. J. Wenthold and C. Hunter, "Immunocytochemistry of Glycine and Glycine Receptors in the Central Auditory System," in O. P. Otterson and J. Storm-Mathisen (Eds.), *Glycine Neurotransmission*, Wiley, Chichester, England, 1990.
48. M. A. Zarbin, J. K. Walmsley, and M. J. Kuhar, "Glycine Receptor: Light' Microscopic Autoradiographic Localization with Tritiated Strychnine," *J. Neuroscience*, Vol. 1, 1981, pp. 532–547.
49. D. M. Caspary, "Cochlear Nuclei: Functional Neuropharmacology of the Principal Cell Types," in R. A. Altschuler, D. W. Hoffman, and R. P. Bobbin (Eds.), *Neurobiology of Hearing*, Raven Press, New York, 1986.
50. D. A. Godfrey, K. L. Beranek, L. Carlson, J. A. Parli, J. D. Dunn, and C. D. Ross, "Contributions of Centrifugal Innervation to Choline Acetyltransferase Activity in the Cat Cochlear Nucleus," *Hear. Res.*, Vol. 49, 1990, pp. 259–280.
51. R. J. Wenthold and M. R. Martin, "Neurotransmitters of the Auditory Nerve and Central Auditory System," in C. Berlin (Ed.), *Hearing Science: Recent Advances*, College Hill Press, San Diego, 1984.
52. M. R. Martin and J. C. Adams, "Effects of DL-*a*-Aminoadipate on Synaptically and Chemically Evoked Excitation of Anteroventral Cochlear Nucleus Neurons of the Cat," *Neuroscience*, Vol. 4, 1979, pp. 1097–1105.
53. D. Oertel, "Cells in the Anteroventral Cochlear Nucleus Are Insensitive to Glutamate and *L*-aspartate; Excitatory Synaptic Responses Are Not Blocked by Aminoadipate," *Brain Res.*, Vol. 302, 1984, pp. 213–220.
54. S. J. Potashner, N. Lindberg, and D. K. Morest, "Uptake and Release of GABA in the Guinea Pig Cochlear Nucleus after Axotomy of Cochlear and Centrifugal Fibers," *J. Neurochem.*, Vol. 15, 1985, pp. 1558–1566.

113

PARALLEL-HIERARCHICAL PROCESSING OF COMPLEX SOUNDS FOR SPECIALIZED AUDITORY FUNCTION

NOBUO SUGA

1 INTRODUCTION

In auditory neuroethology, which is based on the philosophy that the vocal and auditory systems have evolved together for acoustic communication or echolocation, biologically important sounds have been used as stimuli to explore auditory mechanisms underlying species-specific behavior. Such studies on the auditory systems of different types of animals have been successful in finding the neural mechanisms that are species specific as well as the mechanisms that are shared by different types of animals. In this chapter, the neuroethological data obtained from an "auditorily specialized" animal is explained because these will contribute most to the understanding of neural mechanisms for processing complex sounds, including speech sounds. Speech sounds are much more complex than animal sounds, but both types of sounds contain similar basic components, that is, information-bearing elements. These elements and the relationships among them are characterized by particular values of parameters, that is, information-bearing parameters. The central neural mechanisms shared by different types of animals are the production of neurons tuned to IBPs, complex-sound processing by combination-sensitive neurons, and anatomical parcellation for representation of IBPs. Since the best neuroethological data on complex-sound processing were obtained from the mustached bat, this chapter represents the essential portions of these data.

2 INFORMATION-BEARING ELEMENTS AND PARAMETERS IN BIOLOGICAL SOUNDS

In higher vertebrates, communication sounds are usually complex. The amplitude spectra of these sounds commonly change with time. Human speech consists of various phonemes combined in different sequences and is consequently complicated. Sound spectrograms of animal sounds, including speech sounds, exhibit three basic components (information-bearing elements: IBEs): constant frequency (CF), noise burst (NB), and frequency-modulated (FM) components. For instance, the sound spectrogram of a consonant consists of a vertical bar (fill or NB component), indicating the scatter of sound energy over many frequencies. The sound spectrogram of a vowel consists of several horizontal bars (formants or CF components). When two phonemes are combined to form a monosyllabic word, a new component appears as an oblique bar (transition or FM component) (see Part XIII).

Auditory information is carried not only by the acoustic parameters characterizing each of the above three types of IBEs but also by parameters representing relationships among these IBEs in the frequency, amplitude, and time domains. Values of a parameter comprise a continuum. However, only a limited portion of the continuum is important for each species. This limited portion of the continuum has been called the information-bearing parameter (IBP).

The three types of IBEs are found in sounds produced by many different types of animals. For example, the mustached bat, *Pteronotus parnellii*, emits complex biosonar signals consisting of CF and FM components. These are IBEs carrying different types of biosonar infor-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

mation. It also emits a variety of communication sounds that contain three types of IBEs. These sounds differ from biosonar signals in amplitude spectrum and duration. The central auditory system of different species of animals contains not only neurons selectively responding to one of the IBEs but also neurons tuned to particular combinations of IBEs. In the mustached bat, it has been demonstrated that arrays of *combination-sensitive* neurons tuned to different values of IBPs are systematically arranged in the auditory cortex and thus form computational maps. Because of space limitation, this chapter mainly describes processing of auditory information by combination-sensitive neurons in the auditory system of the mustached bat.[1,2] (See Refs. 3 and 4 for the data obtained from the other species of animals.)

3 BIOSONAR SIGNAL

For capture of prey (flying insects) and orientation, the mustached bat emits orientation sounds (biosonar pulses), each of which consists of a long constant-frequency (CF) component followed by a short frequency-modulated (FM) component. Since each pulse contains four harmonics (H_{1-4}), there are eight components that can be defined (CF_{1-4}, FM_{1-4} in Fig. 1*b*). In the emitted pulse the second harmonic (H_2) is always predominant, and the frequency of CF_2 is about 61 kHz. The frequency of the CF component is slightly different among individual bats and between sexes. Accordingly, tonotopic representation in the auditory cortex is sexually dimorphic. In FM_2, frequency sweeps down from 61 to

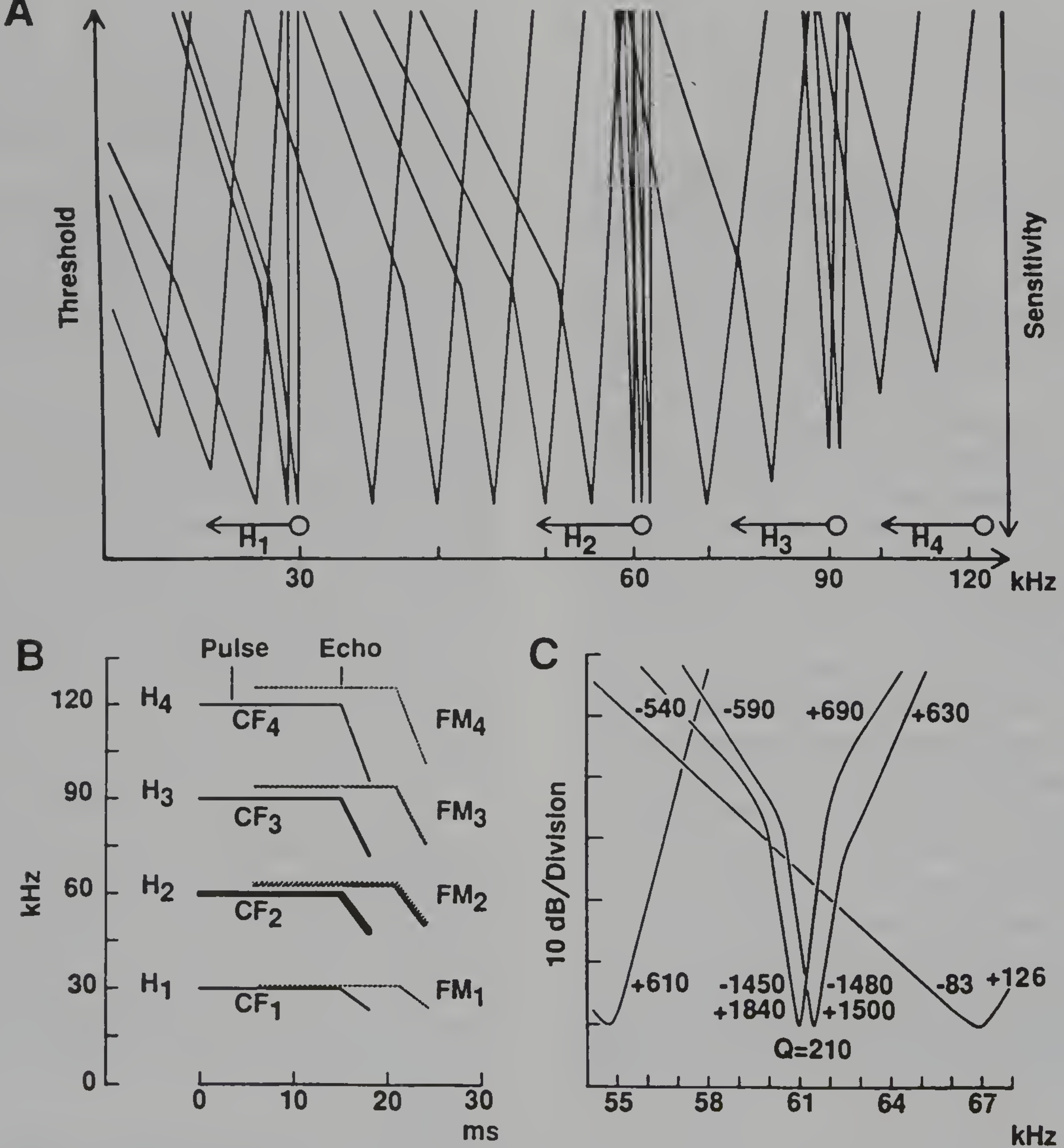


Fig. 1 Biosonar signals and frequency-tuning curves of peripheral neurons of the mustached bat. (*a*) Schematized frequency-tuning curves of 19 peripheral neurons. (*b*) Schematized sound spectrograms of a biosonar pulse and its Doppler-shifted echo. (*c*) Averaged frequency-tuning curves of peripheral neurons tuned to either 54.8, 61.0, 61.5, or 67.0 kHz. The numbers in (*c*) indicate the slopes of the curves in decibels/octave. (Based on N. Suga and Jen, *J. Exp. Biol.*, Vol. 69, 1977, pp. 207–232.)

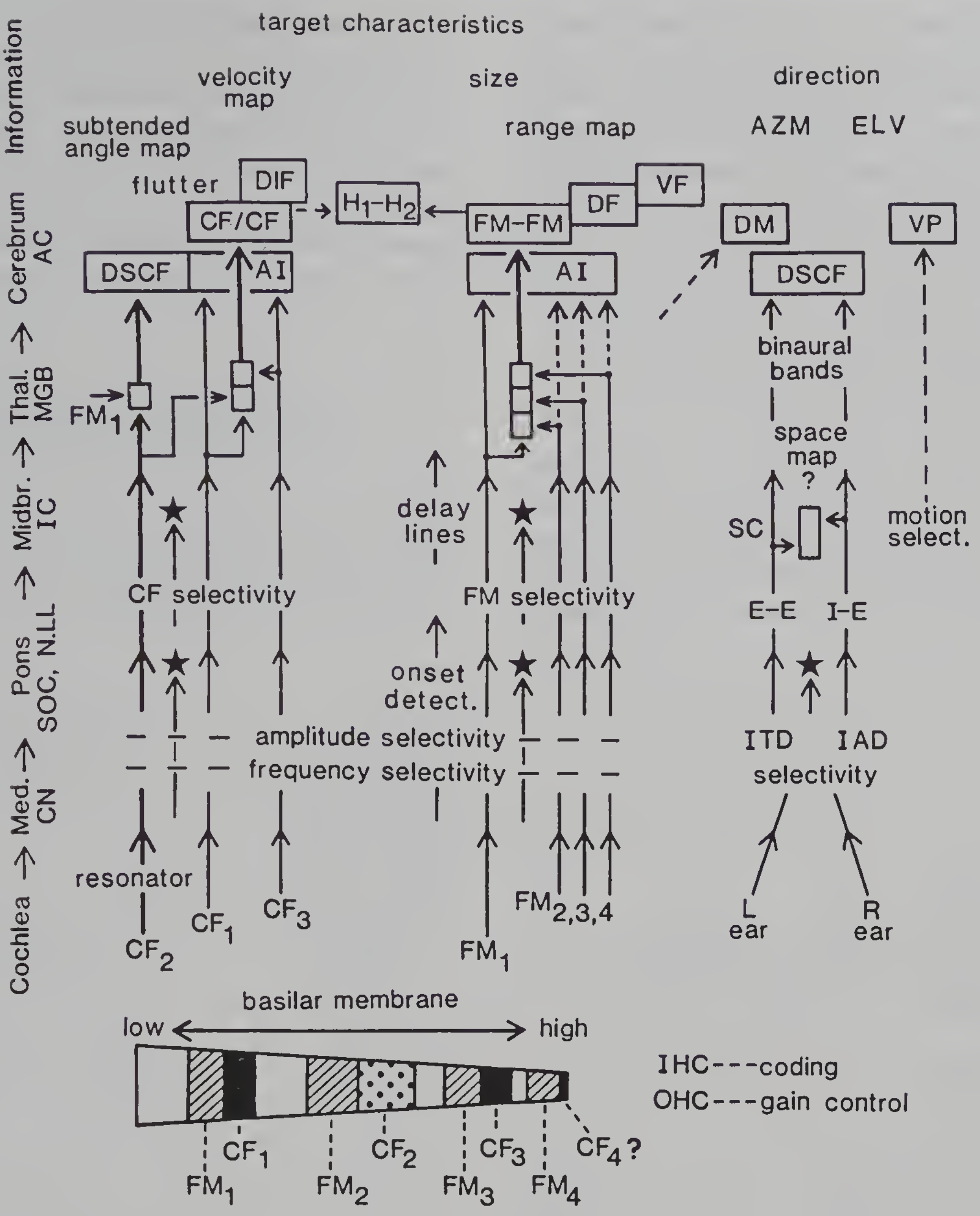

Cochlea → Med. CN → Pons SOC, N.LL → Midbr. IC → Thal. MGB → Cerebrum AC Information
target characteristics
velocity map
size
direction
subtended angle map
range map
AZM ELV
flutter
DIF
CF/CF
H1-H2
FM-FM
DF
VF
DM
VP
DSCF
AI
AI
DSCF
FM1
binaural bands
space map
?
SC
motion select.
delay lines
CF selectivity
FM selectivity
E-E I-E
onset detect.
amplitude selectivity
frequency selectivity
ITD IAD selectivity
resonator
CF1
CF3
CF2
FM2,3,4
FM1
L ear
R ear
basilar membrane
low
high
IHC---coding
OHC---gain control
CF1
CF2
CF3
CF4?
FM1
FM2
FM3
FM4

Fig. 2 Parallel-hierarchical processing of different types of biosonar information carried by complex biosonar signals. The CF$_{1-4}$ and FM$_{1-4}$ of the biosonar pulse and echo are analyzed in parallel at different locations on the basilar membrane in the cochlea (bottom). The signal elements coded by auditory nerve fibers are sent up to the auditory cortex (AC) through several auditory nuclei (left margin): cochlear nucleus (CN), superior olivary complex (SOC), nucleus of lateral lemniscus (NLL), inferior colliculus (IC), and medial geniculate body (MGB). During the ascent of the signals, frequency selectivity is increased in some neurons, and amplitude, CF, and FM selectivities are added to some neurons (arrows with a star). Each star indicates that the addition of selectivity also takes place in the auditory nuclei and cortex as well as in the nucleus where the arrow starts. The CF$_2$ channel is associated with a sharply tuned local resonator and is disproportionately large. It projects to the DSCF area of the auditory cortex. In certain portions of the IC and/or MGB, a hierarchical component in signal processing plays an important role. That is, two channels processing different signal processing elements (e.g., CF$_1$ and CF$_2$ or FM$_1$ and FM$_2$) are integrated to produce CF/CF and FM-FM combination-sensitive neurons. These combination-sensitive neurons respectively project to the CF/CF and FM-FM areas of the auditory cortex, where target velocity or range information is systematically represented, that is, mapped. FM-FM neurons act as delay-dependent multipliers. The delay lines utilized by them are presumably created in the inferior colliculus. Because of corticocortical connections in addition to thalamic projections, DF, VF, and VA areas also consist of combination-sensitive neurons (center top). The DSCF area has the frequency vs. amplitude coordinates to represent velocity and subtended angle information of a target. The DSCF area consists of two subdivisions mainly containing IE or EE binaural neurons. Signal processing in the auditory system is parallel-hierarchical (Update of Ref. 1.)

about 49 kHz. The CF components are an ideal signal for target detection and the measurement of target velocity (relative movements and wing beats of insects), while the short FM components are suited for ranging, localizing, and characterizing a target. Different parameters of echoes received by the bat carry different types of information about a target.

4 FREQUENCY ANALYSIS AT PERIPHERY

The eight components (CF_{1-4} and FM_{1-4}) of the pulse all differ from each other in frequency and are therefore analyzed in parallel in different regions of the basilar membrane (Fig. 2, bottom). For simplicity, we may consider that there are eight channels for the processing of these signal elements: CF_1 channel, CF_2 channel, and so on. Peripheral neurons are more sharply tuned in frequency to the CF components than to the FM components (Fig. 1*a*). In other words, the neurons utilized for velocity measurement are specialized for fine frequency analysis. Among the CF channels, the CF_2 channel is unique because it is associated with an extraordinarily sharply tuned local resonator in the cochlea. The slopes of the frequency-tuning curves of single peripheral neurons tuned to 61 kHz are as steep as −1450 and 1850 dB/octave (Fig. 1*c*). The CF_2 channel is also unique in its size. It occupies about one-third of the ascending auditory system, through the auditory cortex. The sharp frequency tuning and the large population of neurons for coding CF_2 are adaptations for fine frequency analysis and discrimination, that is, for the analysis of relative target velocity and wing beats of insects.

However, even in neurons tuned to 61 kHz (Fig. 1*a*

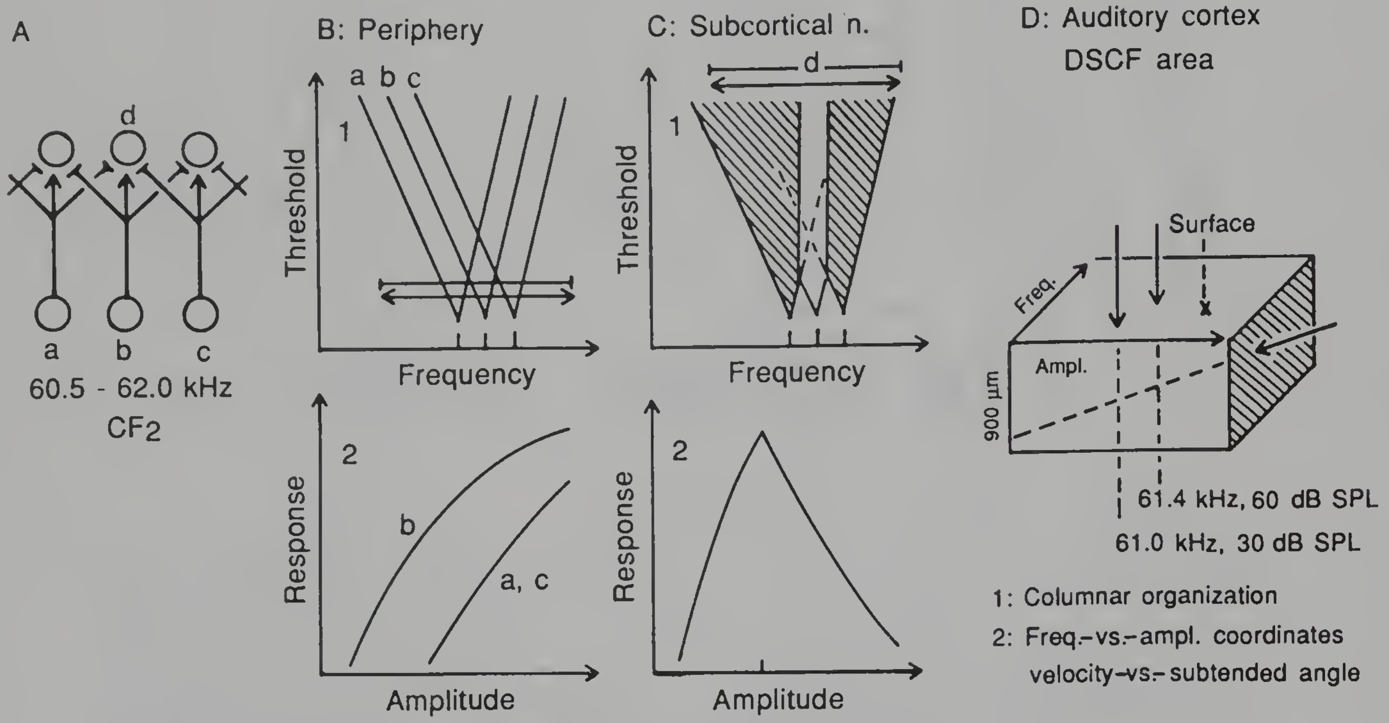


Fig. 3 Sharpening of frequency-tuning curves and creation of amplitude tuning by inhibition. (*a*) Neural circuit for lateral inhibition. The arrows and bar heads represent excitatory and inhibitory synapses, respectively. (*b*) Frequency-tuning curves (*b*1) and impulse-count functions (*b*2) of three peripheral neurons: a, b and c in (*a*). The impulse-count functions are measured with a tone burst at a best frequency of neuron b. The double-headed arrow in (*b*1) indicates that the neurons respond to upward as well as downward sweeping FM sounds and that the threshold of the response is slightly higher than that to a CF tone at a neuron's best frequency. The horizontal bar in (*b*1) indicates that these neurons respond to a noise burst and that the threshold of the response is slightly higher than those to the FM sounds. (*c*) A subcortical neuron [d in (*a*)] has the sharp level-tolerant frequency-tuning curve sandwiched between the inhibitory tuning curves (shaded) as a result of lateral inhibition (*c*1). This subcortical neuron is tuned to a weak stimulus because of the inhibition (*c*2). The double-headed arrow and horizontal bar in (*c*1) indicate that neuron *d* does not respond to FM sounds and noise bursts (Based on N. Suga, in *Basic Mechanisms in Hearing*, A. R. Møller, (Ed.), Academic Press, New York, pp. 675–744. 1973). (*d*) The functional organization of the DSCF area in the auditory cortex. The DSCF area shows columnar organization characterized by a particular combination of best frequency and best amplitude. It has the frequency vs. amplitude coordinates in the plane parallel to the cortical surface (Based on N. Suga, *J. Neurophysiol.*, Vol. 47, 1982, pp. 225–255.)

and *c*), frequency-tuning curves of peripheral neurons are similar to an inverted triangle, so that there is considerable ambiguity in coding frequency by a single neuron at high stimulus amplitudes. All peripheral neurons respond not only to tone bursts but also to FM sounds and noise bursts. (Fig. 3*b*, top). They are not at all selective in repsonding to a particular type of sound. They are insensitive to the amplitude spectrum.

5 PROCESSING OF FREQUENCY AND AMPLITUDE IN CENTRAL AUDITORY SYSTEM

In the CF_1, CF_2, and CF_3 channels at levels higher than the auditory nerve, frequency selectivity of many neurons is increased by lateral inhibition (Fig. 2). Lateral inhibition does not sharpen the "tip 10-dB" portion of frequency tuning curves, but their tail portion, so that the ambiguity in frequency coding remains small even at high stimulus amplitudes.[5] Neurons with such "level-tolerant" frequency tuning curves sandwiched between inhibitory tuning curves respond to tone bursts (CF tones) at particular frequencies, but neither to FM sounds nor noise bursts (Fig. 3*c*, top). The extent of neural sharpening in frequency tuning is most dramatic in the CF_2 channel, so that many neurons in this channel selectively respond to CF tones in a very narrow frequency range regardless of amplitude.

At the periphery, auditory neurons show a monotopic impulse-count function: the stronger the stimulus, the larger the number of impulses discharged (Fig. 3*b*, bottom). They are not tuned in amplitude. Many neurons in the central auditory system, however, are tuned to particular amplitudes: "best" amplitudes (Fig. 3*c*, bottom). Such amplitude tuning or selectivity is created by inhibition occurring at larger stimulus amplitudes. The creation of amplitude selectivity is prominent in the CF_2 channel compared with other CF channels. Thus, the great majority of neurons in the CF_2 channel are tuned both in frequency and amplitude.

As described above, the CF_2 channel is unique. It projects to the Doppler-shifted CF processing (DSCF) area in the primary auditory cortex through the ventral division of the medial geniculate body (MGBv) (Fig. 2). The DSCF area is very large, and neurons in a single column have identical best frequencies and best amplitudes. In other words, each column is characterized by a particular combination of best frequency and best amplitude. Along the cortical surface of the DSCF area, both best frequency and best amplitude systematically vary, forming the frequency vs. amplitude coordinates. That is, these are mapped (Figs. 3*d* and 4*b*). The frequency axis is radial, representing 60.6–62.3 kHz. The best frequencies of single neurons often change at a rate of approximately 20 Hz/column at a certain portion of the DSCF area. The amplitude axis is circular, representing 13–98 dB SPL. Inactivation experiments with muscimol indicate that the DSCF area is necessary for fine frequency analysis at around 61 kHz.

The CF_1 and CF_3 channels, respectively, project to the primary auditory cortex posterior and anterior to the DSCF area. In these areas, tonotopic representation is systematic as in the cat's primary auditory cortex, and, unlike the DSCF area, amplitopic representation has not been found. Thus, the DSCF area is unique both in its very fine tonotopic (echo frequency) representation and amplitopic (subtended target angle) representation.

In addition to the frequency vs. amplitude coordinates, the DSCF area has "binaural" bands or subdivisions that are mainly occupied by either E–E or I–E neurons (Fig. 2, right column and Fig. 4*b*). In the DSCF area, about 50% of the neurons are sensitive to periodic frequency modulation that would be found in echoes from flying insects, but the remaining 50% are not. Furthermore, our recent studies indicate that DSCF neurons show facilitative responses to combinations of H_1 and CF_2. Thus, the functional organization of the DSCF area may be more complex than we now know.

6 EXTRACTION OF VELOCITY INFORMATION IN CENTRAL AUDITORY SYSTEM

To measure relative target velocity, the frequencies of echo CF components must be compared with those of pulse CF components. The central auditory system creates sharply tuned, level-tolerant neurons by lateral inhibition, perhaps at subthalamic nuclei. In the dorsal division of the medial geniculate body (MGBd), part of the CF_1 channel and part of either the CF_2 or CF_3 channel are integrated, so that neurons in this region respond poorly to the CF_1, CF_2, and CF_3 tones when delivered alone, but respond strongly when the CF_1 tone is delivered together with either the CF_2 or the CF_3 tone. As a result of this "hierarchical" component in signal processing, two major types of CF/CF combination-sensitive neurons are created: CF_1/CF_2 and CF_1/CF_3. These CF/CF neurons project separately to the CF/CF area in the auditory cortex and form the CF_1/CF_2 and CF_1/CF_3 subdivision (Figs. 4 and 5).

Each CF/CF neuron has two frequency-tuning curves for facilitation: One is tuned to CF_1 (~29 kHz), and the other is tuned to CF_2 (~61 kHz) of CF_3 (~92 kHz). (See Fig. 5.) A critical parameter for the facilitation of almost all CF/CF neurons is a deviation of the CF_2 or CF_3 frequency from the exact harmonic relationship with the CF_1 frequency, that is, an amount of Doppler shift in terms of CF_2 or CF_3. Some CF/CF neurons, however, are tuned to a combination of two tones that are exact harmonics. The facilitative frequency-tuning curves (facili-

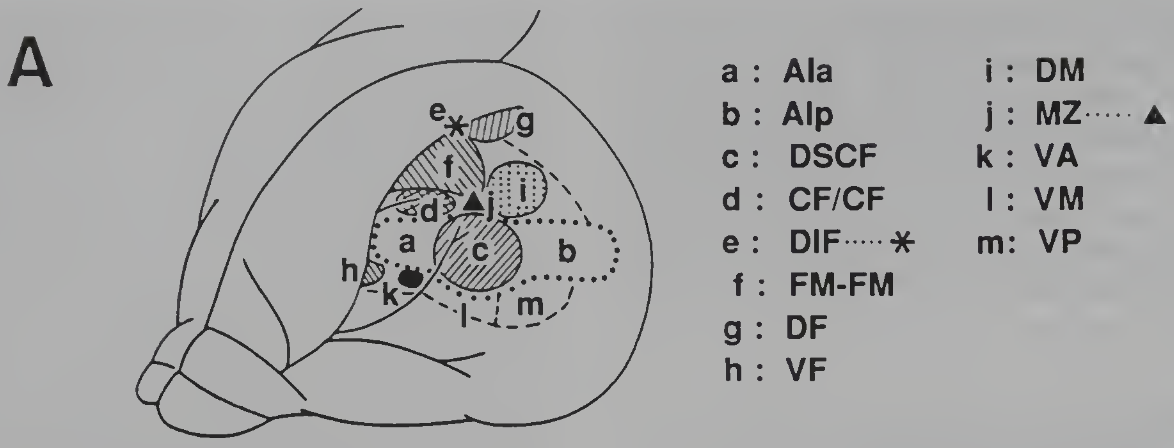


Fig. 4 Functional organization of the auditory cortex of the mustached bat. (*a*) Dorsolateral view of the left cerebral hemisphere. The auditory cortex consists of several areas (a–m). DSCF, CF/CF, FM-FM, DF, VF, and DM areas (c, d, f, g, h, and i, respectively) are specialized for the systematic representation of biosonar information. The branches of the median cerebral artery are shown by the branching lines. The longest branch is on the fossa. (*b*) Graphic summary of the functional organization of the auditory cortex. The tonotopic representation of the primary auditory cortext (the area surrounded by a dotted line in zone d) and the functional organization of the DSCF, CF/CF, FM-FM, DF, and DM areas are indicated by lines and arrows. The DSCF area has axes representing either target velocity (echo frequency: 60.6–62.3 kHz) or subtended target angle (echo amplitude: 13–98 dB SPL) and is divided into two subdivisions suitable for either target detection (shaded) or target localization (unshaded). These subdivisions are occupied mainly by excitatory–excitatory (E–E) or inhibitory–excitatory (I–E) binaural neurons, respectively. The FM-FM area consists of three major types of FM-FM combination-sensitive neurons ($\mathrm{FM}\text{-}\mathrm{FM}_2$, $\mathrm{FM}\text{-}\mathrm{FM}_3$, and $\mathrm{FM}_1\text{-}\mathrm{FM}_4$), which form separate clusters. Each cluster has an axis representing target ranges from 7 to 310 cm (echo delay: 0.4–18 ms). The dorsoventral axis of the FM-FM area probably represents fine target characteristics. The CF/CF area consists of two major types of CF/CF combination-sensitive neurons ($\mathrm{CF}_1/\mathrm{CF}_2$ and $\mathrm{CF}_1/\mathrm{CF}_3$), which aggregate in independent clusters. Each cluster has two frequency axes and represents target velocities from −2 to +9 m/s. The DF and VF areas consist of three types of FM-FM neurons and have a delay axis up to 9 and 5 ms, respectively. The VA area contains only H_1-H_2 combination-sensitive neurons. The DM area has an azimuthal axis representing contralateral 4°–45° [Update of N. Suga, in G. M. Edelman et al. (Eds.), *Dynamic Aspects of Neocortical Function*, Wiley, New York, 1984, pp. 315–373.]

atative areas) of CF/CF neurons are very sharp, level-tolerant, and sandwiched between inhibitory areas. Therefore, these CF/CF neurons are specialized in responding to particular combinations of two frequencies, regardless of stimulus amplitudes. The width of the facilitative frequency-tuning curves of some neurons in the CF_2 and CF_3 channels are only ±1.0% of the best frequencies regardless of stimulus amplitudes. They do not respond to the combinations of FM sounds in the pulse-echo pairs, and probably not to noise bursts, because these would stimulate not only the facilitative but also the inhibitory areas.

In the CF/CF area, neurons tuned to particular combinations of two CF tones are arranged orthogonally to the cortical surface. In other words, each location along the cortical surface represents a particular combination of two frequencies. The combination of two frequencies systematically varies along the cortical surface, so that there are frequency vs. frequency coordinates. Within these coordinates, amounts of Doppler shifts in terms of CF_2 or CF_3, that is, relative target velocities are systematically represented, that is, mapped. This map represents relative target velocities from 8.6 to −1.2 m/s (Figs. 4*b* and 5*d*). CF/CF neurons show sharp *level-tolerant* frequency tuning and are remarkably specialized to respond to a particular frequency relationship between the two CF tones, as a result of the *parallel-hierarchical* processing of complex sounds. (It is explained at the end why velocity information is processed by $\mathrm{CF}_1/\mathrm{CF}_2$ and $\mathrm{CF}_1/\mathrm{CF}_3$ neurons).

7 EXTRACTION OF DISTANCE INFORMATION IN CENTRAL AUDITORY SYSTEM

For echolocation, the bat emits a biosonar pulse and listens to an echo. The echo delay is a primary cue for measuring target range. A 1.0-ms echo delay corresponds

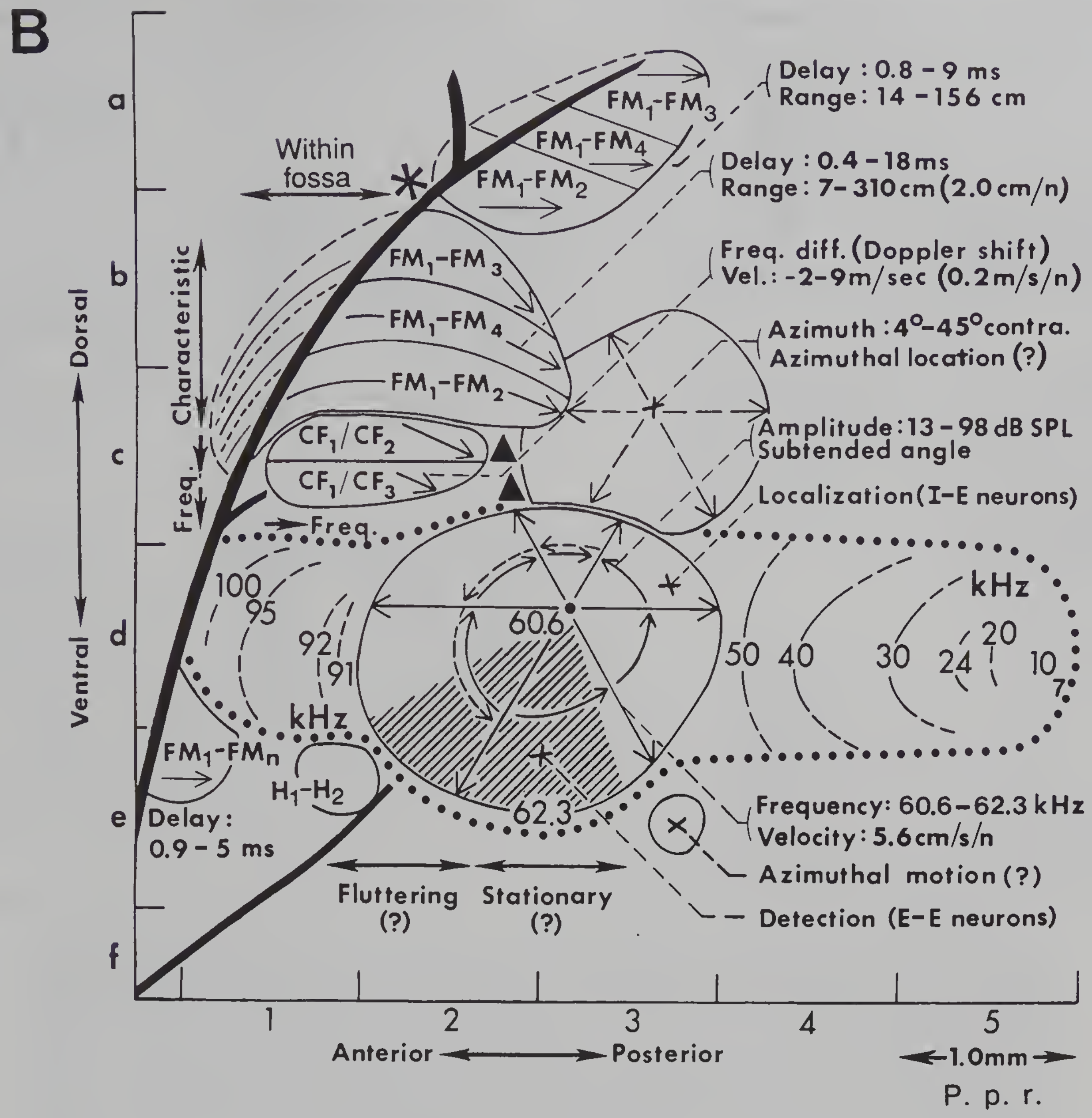

B
a
b
c
d
e
f
Dorsal
Ventral
Characteristic Freq.
Within fossa
FM1-FM3
FM1-FM4
FM1-FM2
Delay : 0.8 – 9 ms
Range: 14 – 156 cm
Delay : 0.4 – 18 ms
Range: 7 – 310 cm (2.0 cm/n)
Freq. diff. (Doppler shift)
Vel.: -2–9 m/sec (0.2 m/s/n)
Azimuth : 4°–45° contra.
Azimuthal location (?)
Amplitude: 13 – 98 dB SPL
Subtended angle
Localization (I-E neurons)
CF1/CF2
CF1/CF3
Freq.
100
95
92
91
kHz
60.6
62.3
50
40
30
24
20
10
7
FM1-FMn
H1-H2
Delay:
0.9 – 5 ms
Frequency: 60.6 – 62.3 kHz
Velocity: 5.6 cm/s/n
Azimuthal motion (?)
Detection (E-E neurons)
Fluttering (?)
Stationary (?)
1
2
3
4
5
Anterior
Posterior
1.0 mm
P. p. r.

Fig. 4 (*Continued*)

to a target distance of 17.3 cm at 25°C. At the periphery, neurons respond to both the pulse and the echo. The time interval between the two grouped discharges is directly related to the echo delay. In the central auditory system, the FM_{1-4} channels converge to create FM-FM combination-sensitive neurons that are tuned to particular echo delays, that is, target distances, and create echo-delay maps, that is, distance maps.

In the FM_1, FM_2, FM_3, and FM_4 channels, inhibition increases frequency selectivity of some neurons and adds amplitude selectivity to some neurons at higher levels (Fig. 2). The extent of sharpening of frequency tuning is much less in the FM channels than in the CF channels. Interestingly, FM selectivity is also added to some neurons by disinhibition or facilitation, so that these *FM-specialized* neurons selectively respond to FM sounds. Delay lines to spread neural responses over 18 ms are created in the FM_1 channel, perhaps, in the inferior colliculus (IC) (Fig. 6). After these parallel processing of FM signals, hierarchical signal processing takes place. That is, in the IC, part of the FM_1 channel and part of either the FM_2, FM_3, or FM_4 channel are integrated to create three major types of FM-FM combination-sensitive neurons: FM_1-FM_2, FM_1-FM_3, and FM_1-FM_4. These collicular FM-FM neurons project to the MGBd in the thalamus. The response properties of thalamic FM-FM neurons are improved in the MGBd, so that these thalamic neurons respond poorly to single FM sounds, but respond strongly to the FM_1 sound combined with either the FM_2, FM_3, or FM_4 sound. The delay of the FM_2, FM_3, or FM_4 sound from the FM_1 sound (hereafter, echo delay) is the critical parameter for their facilitative responses. These FM-FM neurons act as *delay-dependent multipliers* for processing target range information.

The three types of FM-FM neurons in the MGBd separately project to the FM-FM area of the auditory cor-

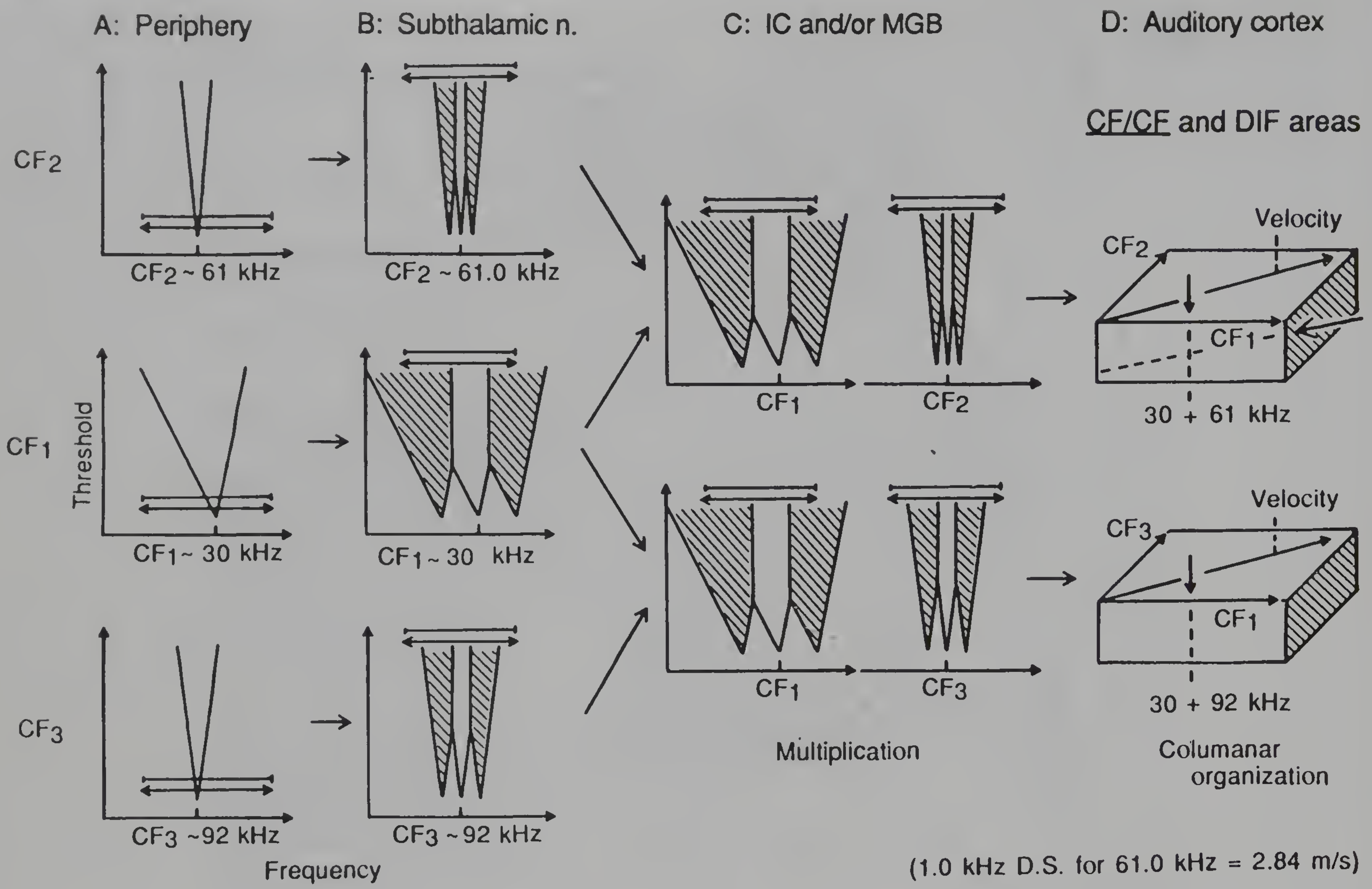

A: Periphery
B: Subthalamic n.
C: IC and/or MGB
D: Auditory cortex
CF/CF and DIF areas
CF2
CF1
CF3
Threshold
CF2 ~ 61 kHz
CF2 ~ 61.0 kHz
CF1 ~ 30 kHz
CF1 ~ 30 kHz
CF3 ~92 kHz
CF3 ~ 92 kHz
Frequency
CF1
CF2
CF1
CF3
Multiplication
Velocity
CF2
CF1
30 + 61 kHz
Velocity
CF3
CF1
30 + 92 kHz
Columanar organization
(1.0 kHz D.S. for 61.0 kHz = 2.84 m/s)

Fig. 5 Neural mechanisms for creating the frequency vs. frequency coordinates. Triangular frequency-tuning curves (*a*) at the periphery are changed into level-tolerant frequency-tuning curves (*b*) in subthalamic auditory nuclei by lateral inhibition. Then, the CF_1 and CF_2 or CF_3 channels are integrated in the inferior colliculus and/or the medial geniculate body, creating two types of CF/CF combination-sensitive neurons (*c*). These CF/CF neurons project to the CF/CF area in the auditory cortex and form the CF_1/CF_2 and CF_1/CF_3 subdivisions. In each subdivision, a cortical column is characterized by a particular combination of two frequencies. Each subdivision has frequency vs. frequency coordinates, in the plane parallel to the cortical surface. In the coordinates, target-relative velocities are systematically represented, that is, mapped (Update of Ref. 1.)

tex, forming FM_1-FM_2, FM_1-FM_3, and FM_1-FM_4 subdivisions. In each subdivision, neurons tuned to particular echo delays are arranged orthogonally to a cortical surface. In other words, each location along the cortical surface represents a particular echo delay. The echo delay represented by neurons (best delay) systematically varies along the cortical surface. Thus, FM-FM neurons form an echo-delay axis in each subdivision for the systematic representation of target range information (Fig. 6). The FM-FM area represents target ranges from 7 to 310 cm (Fig. 4*b*). The signal processing in the FM channels is parallel-hierarchical. (It is explained later why distance information is processed by FM_1-$FM_{2, 3, \text{or} 4}$ neurons).

The FM-FM area projects to the DF and VF areas of the auditory cortex in addition to other regions of the brain. The DF area consists of three subdivisions containing FM-FM neurons. In each subdivision, target ranges of up to 160 cm are systematically represented (Fig. 4). The DF area projects to the VF area as well as other areas of the brain. The VF area also consists of three clusters of FM-FM neurons and represents target ranges of up to 80 cm. The functional significance of these multiple-range (delay) axes are presumably related to biosonar behavior at different distances to targets.

As described above, part of one channel is integrated with part of another channel in the IC and/or MGB. The remaining parts of these channels project to the auditory cortex, which is not described above. The auditory cortex of the mustached bat thus shows multiple cochleotopic (tonotopic) representations, which are related to the representations of different types of biosonar information (Fig. 4). In each cortical area, certain responses proper-

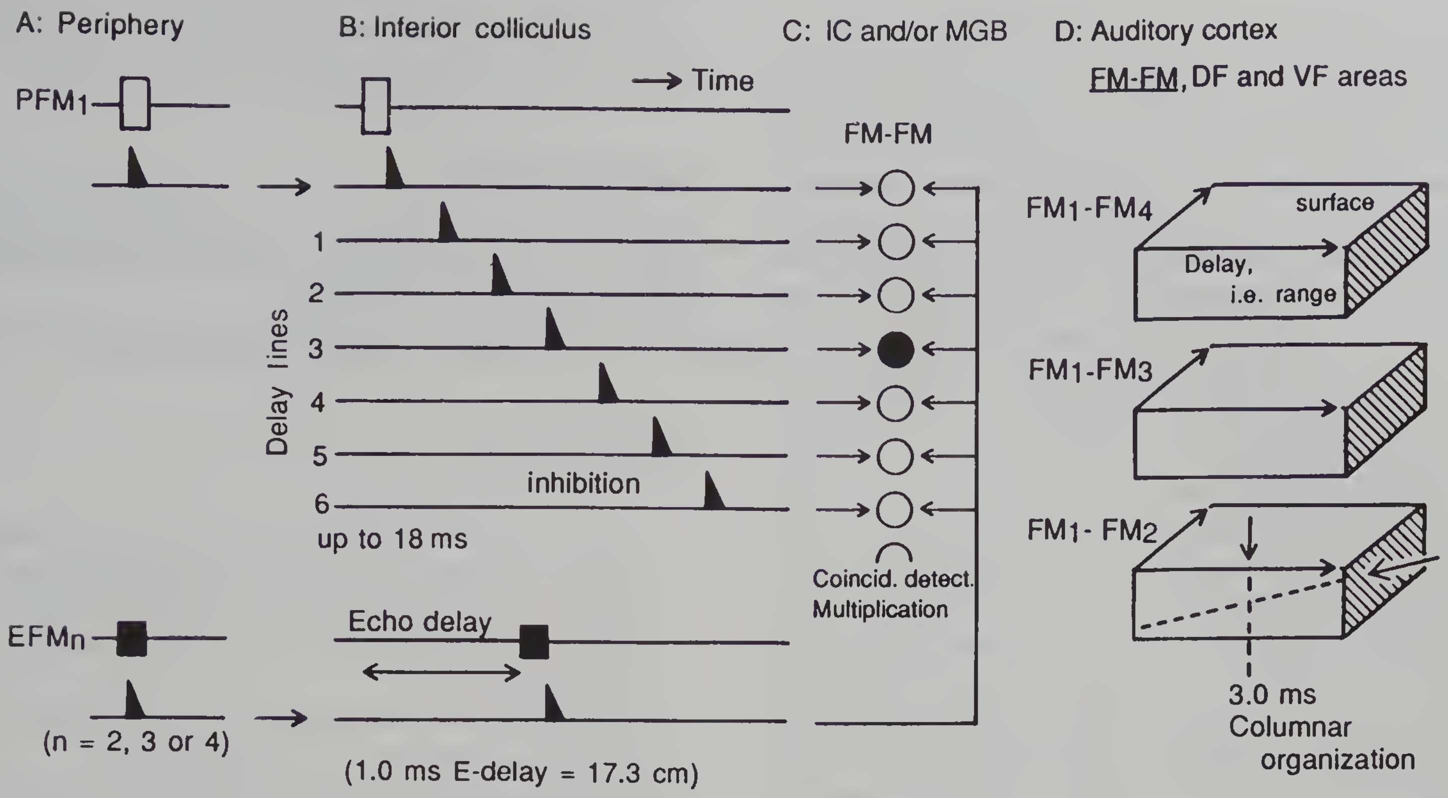


Fig. 6 Neural mechanisms for creating delay (target-range)-tuned neurons. In the inferior colliculus (IC), each iso-best frequency slab has a delay line. Delay-tuned FM-FM neurons in the IC have two inputs: an input to receive signals from neurons tuned to frequencies swept by the FM_1 (FM_1 channel), which are associated with delay lines from 0.4 to 18 ms, and an input to receive signals from neurons tuned to frequencies swept by FM_n (FM_n channel: $n = 2, 3$, or 4), which are not associated with delay lines (i.e., neurons associated only with a shortest portion of the delay lines). For example, the FM-FM neuron indicated by a filled circle in (*c*) receives signals from an FM_1 neuron associated with a 3.0-ms delay line and also from an FM_n neuron associated with 0 ms delay line. When this neuron receives a signal from the FM_n channel activated by an echo FM_n, which delays 3.0 ms from the emitted pulse FM_1, this neuron receives the signals from the FM_1 and FM_n channels at the same time. The amount of coincidence of these two signals is amplified at the synapse of this FM-FM neuron. Three types of thalamic FM-FM neurons separately project to the FM-FM area of the auditory cortex, forming three subdivisions within the FM-FM area (*d*). Each subdivision of the FM-FM area shows columnar organization characterized with a particular value of echo delay. It has an echo delay axis for the systematic representation of target distances on the plane parallel to the cortical surface. (Based on Ref. 2.)

ties of single neurons arranged orthogonally to the cortical surface are identical. Along the cortical surface, however, the response properties vary systematically, so that each area has an axis for the representation of a particular type of biosonar information. Among the various functional areas, the CF/CF, FM-FM, DF, VF, H_1-H_2 (part of the VA) and DSCF areas contain combination-sensitive neurons, so that these areas are particularly interesting in terms of neural mechanisms for processing complex sounds. The H_1-H_2 area contains combination-sensitive neurons that are different from FM-FM and CF/CF neurons. They show facilitative responses to the CF_2 and/or FM_2 of an echo when these are combined with the CF_1 and/or FM_1 of the biosonar pulse. They are broadly tuned to echo delays and unnaturally high velocities. Neurons in the DSCF area are sensitive to combinations of pulse FM_1 and echo CF_2. They remain to be further studied.

The projections of the CF/CF area may overlap with those of the FM-FM area in several regions of the brain. The physiological data thus far obtained indicate that complex acoustic signals are processed in a parallel-hierarchical way in the ascending auditory system and beyond the auditory cortex.

8 PROTECTION OF NEURAL COMPUTATION FROM MASKING

CF/CF and FM-FM neurons extract velocity or distance information by comparing the higher harmonics of the Doppler-shifted echo with the first harmonic of an emitted pulse. The first harmonic (CF_1 and FM_1) is the weakest component. It has less than 1% of the total energy of the emitted pulse. However, it is used as the reference to measure target relative velocity and distance. Why does the auditory system create such *heteroharmonic-sensitive* neurons instead of *homoharmonic-sensitive* neurons? This heteroharmonic sensitivity is one of the mechanisms used to protect the neural processing of biosonar information from masking by the sounds produced by conspecifics. Let us consider that hundreds of bats are flying in a cave and that the first harmonic of their pulses is completely suppressed by the antiresonance of the vocal tract. Then in the air, there are many second, third, and fourth harmonics produced by conspecifics. The combinations of these harmonics in the sounds produced by the conspecifics cannot excite the heteroharmonic-sensitive FM-FM and CF/CF neurons because the first harmonic is absent. When the animal itself emits pulses, the first harmonic is present at the larynx and stimulates both the ears by bone conduction. The FM-FM and CF/CF neurons are then turned on to process target distance or velocity information. In this way, the neural computation of biosonar information in the auditory cortex is protected from the masking that would be caused by the sounds emitted by many conspecifics flying nearby.

Note: After writing this Chapter, two major findings have been made: (1) the precursors of cortical combination-sensitive neurons are in the inferior colliculus, and (2) the corticofugal system adjusts/improves signal processing in the frequency and time domains.

Acknowledgment

This work was supported by NIDCD research grant DC00175, ONR research grant N00014-90-J-1068, and The McKnight Foundation.

REFERENCES

1. N. Suga, "Auditory Neuroethology and Speech Processing: Complex Sound Processing by Combination-Sensitive Neurons," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function*, Wiley, New York, 1988, pp. 679–720.
2. N. Suga, "Cortical Computational Maps for Auditory Imaging," *Neural Networks*, Vol. 3, 1990, pp. 3–21.
3. A. N. Popper and R. R. Fay, *The Mammalian Auditory Pathway: Neurophysiology*, Springer-Verlag, New York, 1992.
4. A. N. Popper and R. R. Fay, *Hearing by Bats*, Springer-Verlag, New York, 1995.
5. N. Suga and T. Manabe, "Neural Basis of Amplitude-Spectrum Representation in Auditory Cortex of the Mustached Bat," *J. Neurophysical.*, Vol. 47, 1982, pp. 225–255.

PART XII

PSYCHOLOGICAL ACOUSTICS

114

INTRODUCTION

DAVID M. GREEN AND DENNIS MCFADDEN

1 INTRODUCTION

Traditionally, the study of the auditory system has been pursued by scientists of two general types using different experimental subjects and procedures. On the one hand are auditory physiologists, who typically study nonhumans. They use all the modern tools of scientific technology: electrical recordings of various sorts, neural staining, labeling, biochemical tracers, electron microscopy, and so on. Their goal is to better describe the construction, organization, and function of individual elements and subsystems at the various stages of processing in the auditory nervous system.

The other group consists of auditory psychophysicists, or psychoacousticians. They generally study humans, using experimental paradigms and acoustic stimuli that are chosen for their ability to reveal how the human listener senses and perceives auditory stimuli. The methods are behavioral, and the products of this research include numerical scales of the major auditory attributes and measures of the limits of auditory performance as well as the effects of various stimulus parameters on those limits. The aim of psychological acoustics is to provide a functional description of the hearing process. The approach is to treat the auditory system as an information-processing system and to describe the salient features of that system. The chapters in this section describe the current status of knowledge in several representative topic areas of psychological acoustics.

One of the purposes of introductory chapters of this kind is to inform curious readers from other disciplines of what might be of interest to them in the chapters that follow. Our presumption is that many readers of the following section will be engineers and other physical scientists who are generally knowledgeable about acoustics but somewhat unsure about the scientific questions or issues that are addressed by psychological acousticians. Accordingly, this introduction begins with some background information and then discusses some of the specific topics taken up in the following chapters.

2 HUMAN AUDITORY SYSTEM

The human auditory system is impressive for a number of reasons, but two outstanding features are its absolute sensitivity and its enormous dynamic range. Normal listeners can detect sounds that produce an average displacement of the eardrum smaller than the diameter of a hydrogen atom. They can also hear sounds six orders of magnitude larger than this displacement without the familiar, protracted dark and light adaptation that characterizes the visual system. Listeners are also reasonably sensitive to relative change over this entire range; they can discriminate changes in sound amplitude of less than 10%. The auditory system consists of a mechanically sophisticated "front end," the middle ear and cochlea, which transforms variations in air pressure into minute displacements of the basilar membrane. There is increasing evidence that the exquisite sensitivity of the system arises from certain nonlinearities that are present in the basilar membrane. These displacements produce graded potentials in the transducer elements, the hair cells, which in turn lead to neural impulses in the individual neurons constituting the VIIIth cranial nerve. These neurons all terminate (synapse) in the first way-station in the brainstem, the cochlear nucleus. The basilar mem-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

brane is only about 3.5 cm in length and contains about 12,000–16,000 of the sensory receptors called hair cells. The hair cells are arranged in rows, four cells abreast, each inner hair cell flanked by three outer hair cells. It is believed that the approximately 4000 inner hair cells are responsible for most of the information processing that takes place in the cochlea, because about 90% of the afferent neurons make contact with these inner hair cells.

A great deal is also known about the coding of the neural impulses traveling in the VIIIth nerve and about the complexity of neural interaction in the cochlear nucleus. Our understanding of how information is coded grows increasingly more sketchy as we move up the brainstem to the central regions of the auditory system. For example, much less is known about the processing in the auditory cortex than in the visual cortex.

3 PSYCHOACOUSTICAL STUDIES

Psychoacoustical studies try to discover the functional significance of this system by presenting controlled stimuli and measuring the reactions of human listeners. Quantifying these reactions is a matter of some subtlety. One major effort is to use broad categories of responses to develop numerical scales of the major attributes. For example, Chapter 118 describes how scales of loudness are constructed from numerical estimates given by the listeners. Another source of data is a kind of balance judgment—for example, adjusting the frequency of one sound until its pitch is equal to the pitch of some other sound, as described in Chapter 116.

A final source of psychophysical information is the smallest amount of change in some stimulus variable that can be reliably detected or discriminated. For example, how much does one need to change the frequency of a sinusoid in order for listeners to discriminate a difference?

Historically, measures of detection and discrimination were plagued by the problem of uncontrolled response biases. Some subjects—by virtue of their personality characteristics and/or past experience in decision-making situations—would be strongly inclined toward giving one of the two possible responses (signal present or absent, stimuli same or different, etc.) when uncertain about which stimulus condition occurred, as is often the case in experiments designed to measure the limits of sensitivity. The consequence was that classical measures of sensitivity were confounded by response bias. A major advance occurred when it was demonstrated that certain forms of forced-choice experiment had the potential to yield measures of sensitivity that are independent of the subject's response bias or criterion for response. The widespread use of forced-choice procedures in the past 30 years has led to far greater replicability of results in detection and discrimination tasks than was previously true. Modern forced-choice procedures and measures of sensitivity are discussed by Green and Swets,[1] and the results from discrimination experiments comprise much of the material discussed in Chapters 115, 117, and 121. Discriminating the presence or absence of a sinusoidal signal in quiet (the absolute threshold of hearing) is the entire topic of Chapter 123.

These different methods contribute to our understanding of how auditory signals are processed and perceived. Like most biological measurements, there is considerable variability in the measurements. A typical discrimination value, which provides the most precise measurements in the area, often has a standard deviation that is as much as 25% of the mean, even if the measurements are made on the same listener at different times. This variability is even larger when measurements are taken from different individuals. Some investigators believe that understanding these differences between individual listeners will become a major topic of future psychoacoustical research. Certainly there is much to be learned in this area.

Once psychoacoustical data have been collected and analyzed, models or theories are formulated in an attempt to explain the experimental results. These theories are not basically different from scientific theories in other disciplines, with the possible exception that the inferences are often deep—meaning that "black-box" theorizing is common. Psychoacoustical investigators follow with great interest the research of their physiological colleagues, because it often provides insight into the mechanisms that process the stimulus information and convert it to measured responses. Progress is made when a physiological finding helps us understand some psychophysical result, and often there is a gratifying correspondence between the physiological and psychological data.

The past century has witnessed enormous advancements in our understanding of the auditory system. In large part, these advancements stem from our utilization of modern electronic technology. It should be realized that control of the stimulus was virtually impossible in the early part of this century. Controlling the pitch of a sinusoidal sound was comparatively easy, because, in principle, different tuning forks could be constructed to produce vibrations of any desired frequency. But controlling the amplitude of a sinusoidal sound, which is a critical physical parameter in most auditory tasks, was, for the most part, either difficult or impossible. Controlled changes in relative amplitude are possible only in an anechoic room where the inverse-square law operates. Thus, the distance between the source and listener could be used as a means of systematically altering the relative level of two sounds. The measurement of the absolute level of a sound wave was first possible in 1882 with Rayleigh's disc. The refinements of inexpensive earphones and electronics made it possible to achieve convenient control of the two most salient properties of an auditory stimulus—frequency and level—which correspond roughly to pitch and loudness.

4 INTENSITY AND FREQUENCY DISCRIMINATION

The modern era of psychoacoustics was heralded by quantitative studies of two important auditory capacities: intensity discrimination[2] and frequency discrimination.[3] At about the time of these studies, Georg von Bekesy, a Hungarian telephone engineer, began a series of investigations about how the basilar membrane vibrates, for which he was awarded a Nobel prize in 1961. (For an excellent review of that work and a translation of some of his earlier papers, see the work of Bekesy.[4]) Perhaps Bekesy's most interesting discovery was that the basilar membrane supports an unusual form of traveling waves (their amplitude increases and speed decreases with distance from the stapes). From a functional point of view, Bekesy's most important observation was that the place of maximum vibration along the basilar membrane systematically changes with stimulus frequency, as would occur in a resonant system. This confirmed the earlier conjecture of Ohm and Helmholtz that there is a strong tonotopic organization in the cochlea. High-frequency tones stimulate the part of the basilar membrane near the entrance to the canal, whereas low-frequency tones produce their maximum vibrations at the more apical parts.

Another telephone engineer, Harvey Fletcher, working at the Bell Telephone Laboratories in the United States, developed the concept of a critical band, namely, that the first stage in auditory processing can be likened to a set of filters or frequency channels, each tuned to slightly different center frequencies and having a bandwidth of less than a third of an octave.[5,6] Measurements of the detectability of a sinusoidal signal in noises of differing bandwidths suggested that the critical quantity was the ratio of signal power to noise power present in this filter or frequency band. These experiments provided the basic data for this concept, and its applicability to a variety of other psychoacoustical phenomena was soon appreciated. Eberhard Zwicker, in Germany, demonstrated the effects and importance of such filtering in such diverse areas as masking, the ability to discriminate amplitude and frequency modulation, and loudness judgments. Feldtkeller and Zwicker[7] provided a good summary of that research. More recent measurements have led to more precise estimates of the filter shape[8] and the equivalent rectangular bandwidths (ERBs) of these peripheral filters.[9] The importance of this first-stage filter concept is evident in the topics discussed in Chapters 115, 116, and 118.

5 PITCH

While analysis in the frequency domain dominated the early research in hearing, Jan Schouten in Holland began the first of a series of experiments that challenged the ideas of Ohm and Helmholtz as to how the pitch of a complex sound is determined.[10] The primary puzzle was the pitch of the missing fundamental. A periodic sound rich in harmonics is perceived as having a pitch equal to the fundamental, whether or not energy is present at the fundamental frequency. Ohm and Helmholtz argued that the fundamental energy did not need to be present in the stimulus, because the nonlinearity of the transmission process would produce energy at the difference in frequency between the overtones, which is equal to the fundamental frequency. Thus, they could assume that pitch arose from a resonance process causing stimulation of the cochlea at the location tuned to the fundamental frequency of a complex sound. Schouten showed that these ideas were wrong. For example, masking noise can be present at the fundamental frequency (making it inaudible), yet the pitch created by the audible overtones is unchanged. Also, when all components of an harmonic sequence are increased in frequency by the same amount (so the difference frequency remains the same), the pitch of the complex increases slightly. These observations destroyed the rigid connection between place and pitch that was the crux of the Ohm–Helmholtz resonance theory and have led to various theories emphasizing the importance of the periodicity information contained in both the displacement pattern of the basilar membrane and the firing pattern of VIIIth nerve fibers. A review of this fascinating area along with more recent research is the topic of Chapter 108.

A cliché among students of the brain is that the nervous system is primarily concerned with stimuli that change—that is, with stimuli that contain spatial or temporal transients. A corollary of this view is that the brain rapidly adapts to steady-state stimuli. Common examples include the acute awareness of a tight piece of clothing when it is first put on and the obliviousness to it shortly thereafter. The visual system is known for the emphasis it places on edges, boundaries, and onsets. What about the auditory system? While many aspects of loudness, pitch, masking, and so on, involve relatively steady-state stimuli and constant perceptions, there are also situations in which these perceptions can change over time. For example, the detectability of a brief tonal signal can be quite different, depending upon whether it occurs at the beginning, middle, or end of a noise burst. Some of these changes are thought to be related to the dynamic changes in neural firing rate that are observed physiologically when sounds are turned on and off. These adaptation-like phenomena are discussed in Chapter 122.

6 LOCALIZATION OF SOUND

A topic of long-standing interest to many acousticians is how sound is localized in space, an ability that is remarkably acute. Over the years, psychoacoustical research has documented this skill and has examined the nature and range of the acoustical cues underlying it. Nearly 100 years ago, Lord Rayleigh[11] advanced what is known as the duplex theory, namely, that different cues are used in different frequency regions. One set of cues depends on differences in the level of the sound at the two ears (because of the sound shadow thrown by the head); the second cue set depends on differences in time of arrival of the sound at the two ears (because of the slight difference in path length to the further ear). Our ability to use these two cues continues to receive attention. The limitation that Rayleigh presumed to exist on the usefulness of the time cue at high frequencies is now known not to exist when the waveform is complex, as it invariably is in the real world. Our physiological colleagues have established that the cues of interaural time and intensity are processed by different populations of cells at various levels in the auditory system, and they have demonstrated evolutionary trends that link head size, cue effectiveness, and the upper frequency limit of hearing. Other discoveries have revealed that the mechanisms underlying sound localization can also aid in the extraction of certain signals from background noise. In recent times, there has been considerable interest in simulating acoustical environments with headphones or small numbers of loudspeakers (a psychoacoustical virtual reality). Success at this task requires knowledge of the acoustical cues that provide information about the elevations of real-world sound sources as well as their azimuths. These topics, along with the basic facts and theories of sound localization, are discussed in Chapter 117.

7 PERCEPTION OF COMPLEX SPECTRA

Recent psychoacoustical experiments have begun to explore stimuli more complicated than single sinusoids or noise. Data obtained with multitonal complexes, amplitude-modulated noise bands, and other complex spectra have produced results that are inconsistent with the idea that detection is determined solely by the signal-to-noise ratio in a single critical band. In one experiment, Hall, Haggard, and Fernandes[12] showed that Fletcher's original critical-band experiment produced very different results when the entire noise band was amplitude modulated together. As the bandwidth of the noise increased, the threshold actually improved by 10 dB or more. It is now appreciated that the temporal modulation of a spectrally remote noise band can provide a strong cue to the detection of a signal in another noise band if the modulation is correlated in the two bands. In other work, Green[13] has shown that some signals are detected by comparing the relative amplitude of components widely spaced in the spectrum, not just by monitoring the activity in a single frequency channel centered at the signal frequency. From their increasing experience with more complex waveforms, psychoacousticians now realize that the abilities and performance of individual listeners can differ considerably, and the bases for these individual differences are receiving increasing attention. Chapter 121 explores these and other related topics that have expanded our awareness of how more complex waveforms are perceived.

8 IMPAIRED HEARING

Finally, the topic of impaired rather than normal hearing is considered. One important definition of impaired hearing is still based largely on the audibility of single sinusoidal signals presented in quiet—the audiogram. Degradation in speech intelligibility is clearly another important aspect of impaired hearing. Unfortunately, the available data indicate that speech intelligibility correlates with audiometric thresholds only modestly. How the absolute threshold is defined and measured is the topic of Chapter 123. Chapter 120 discusses the different types of hearing impairment and the tests used to determine these categories. It has been clear for some time that exposure to environmenal noise can have profound short- and long-term effects on hearing. While such effects have been recognized for at least a century, quantitative data relating noise exposure and hearing level have only begun to be acquired. Chapter 121 relates what is known and not known in this important area.

The aim of this chapter was to provide the curious outsider with an overview of psychological acoustics and of the material contained in the following chapters. The newcomer to psychoacoustics is encouraged to also study the chapters written by our colleagues in physiological acoustics, for, in our view, the information contained there is crucial to our common, ultimate goal of understanding the behavior of the auditory system. This has always been a joint effort, and it is becoming increasingly so. Pleasant reading.

Lastly, on a more personal note, we wish to thank the authors who wrote chapters for this section. Most were punctual, and all were cooperative and pleasant to deal with. We appreciated that.

REFERENCES

1. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Wiley, New York, 1966; reprinted 1988 Penisular Publishing, Los Altos, CA.
2. R. R. Riesz, "Differential Sensitivity of the Ear for Pure Tones," *Phys. Rev.*, Vol. 31, 1928, pp. 867–875.
3. E. G. Shower and R. Biddulph, "Differential Pitch Sensitivity of the Ear," *J. Acoust. Soc. Am.*, Vol. 3, 1931, pp. 275–287.
4. G. von Bekesy, *Experiments in Hearing*, McGraw-Hill Series in Psychology, translated and edited by E. G. Wever, McGraw-Hill, New York, 1960.
5. H. Fletcher, "Auditory Patterns," *Rev. Modern Phys.*, Vol. 12, 1940, pp. 47–65.
6. H. Fletcher, *Speech and Hearing in Communication*, Van Nostrand, New York, 1953.
7. R. Feldtkeller and E. Zwicker, *Das Ohr als Nachrichtenempfanger*, S. Hirzel Verlag, Stuttgart, 1956.
8. R. D. Patterson, "Auditory Filter Shapes Derived with Noise Stimuli," *J. Acoust. Soc. Am.*, Vol. 59, 1976, pp. 640–646.
9. B. C. J. Moore and B. R. Glasberg, "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," *J. Acoust. Soc. Am.*, Vol. 74, 1983, pp. 750–753.
10. J. F. Schouten, "The Residue Revisited," in R. Plomp and G. F. Smoorenburg (Ed.), *Frequency Analysis and Periodicity Detection in Hearing*, A. W. Sijthoff, Leiden, 1970. An excellent reference to Schouten's early experiments is J. F. Schouten, "Five Articles on the Perception of Sound (1938–1940)," Instituut voor Perceptie Onderzoek, Eindhoven, The Netherlands.
11. Lord Rayleigh, "On Our Perception of Sound Direction," *Philos. Mag.*, Vol. 13, 1907, pp. 214–232.
12. J. W. Hall, III, M. P. Haggard, and M. A. Fernandes, "Detection in Noise by Spectro-temporal Pattern Analysis," *J. Acoust. Soc. Am.*, Vol. 76, 1984, pp. 50–56.
13. D. M. Green, *Profile Analysis: Auditory Intensity Discrimination*, Oxford University Press, New York, 1988.

115

AUDITORY MASKING

SØREN BUUS

1 INTRODUCTION

The purpose of this chapter is to review masking as it applies to hearing. In hearing, masking generally is defined as the interference with the perception of one sound (the signal) by another sound (the masker). The interference may decrease the loudness of the signal, may make a given change in the signal less discriminable, or may make the signal inaudible. The first two cases are often called partial masking. Some cases of partial masking are considered in 80, 118, and 122.

This chapter considers only how a masker changes the audibility of the signal. The change of audibility is usually measured as the *amount of masking*, which is the increase of threshold, in decibels, caused by the presence of the masker. Thus, the amount of masking is defined as the sound pressure level (SPL) at which the signal is just audible in the presence of the masker minus the SPL at which it is just audible in the quiet. However, many aspects of masking are best summarized by the SPL or signal-to-noise ratio at which the signal is just audible. (Here and in the following, "just audible" means that listeners' performance meets some fixed criterion in a detection experiment. The criterion may be defined implicitly or explicitly by the experimental procedure. Whatever criterion is used, however, it should be the same with and without the masker. See Chapter 80 for further discussion of signal detection theory, detectability, and audibility.)

Because quiet thresholds set a lower limit on thresholds obtained under masking and serve as a reference for calculating the amount of masking, they are discussed briefly in the beginning of this chapter (see also Chapter 123). Then follows a discussion of *simultaneous masking*, which is the increase in threshold caused by a masker that is present throughout the duration of the signal. Next, the chapter discusses *nonsimultaneous masking*, which is the increase in threshold caused by a masker presented before the signal (*forward masking*) or after it (*backward masking*). The final section of this chapter considers special aspects of masking, which cannot be explained easily by conventional models of masking. These phenomena include *interaural masking* (also called *central masking*), *comodulation masking release* (CMR), and "*overshoot*." Other phenomena that are not explained by conventional models of masking are *profile analysis* (see Chapter 121) and the *binaural masking level difference* (see Chapter 117).

2 ABSOLUTE THRESHOLDS

Absolute thresholds depend on frequency (see Chapter 123). As discussed below, they also depend on the bandwidth and duration of the signal.

2.1 Effect of Bandwidth

The threshold for a multitone complex with closely spaced components is determined by the total intensity as long as all the components are within some limiting bandwidth.[1] Adding tones outside that bandwidth does not reduce the threshold (in terms of level per tone). As shown in Fig. 1, the limit is about 160 Hz below 1100 Hz and about 320 Hz below 2000 Hz. These bandwidths are closely related to the auditory filter bandwidths, which are discussed in detail later. A recent study replicated these findings.[2]

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

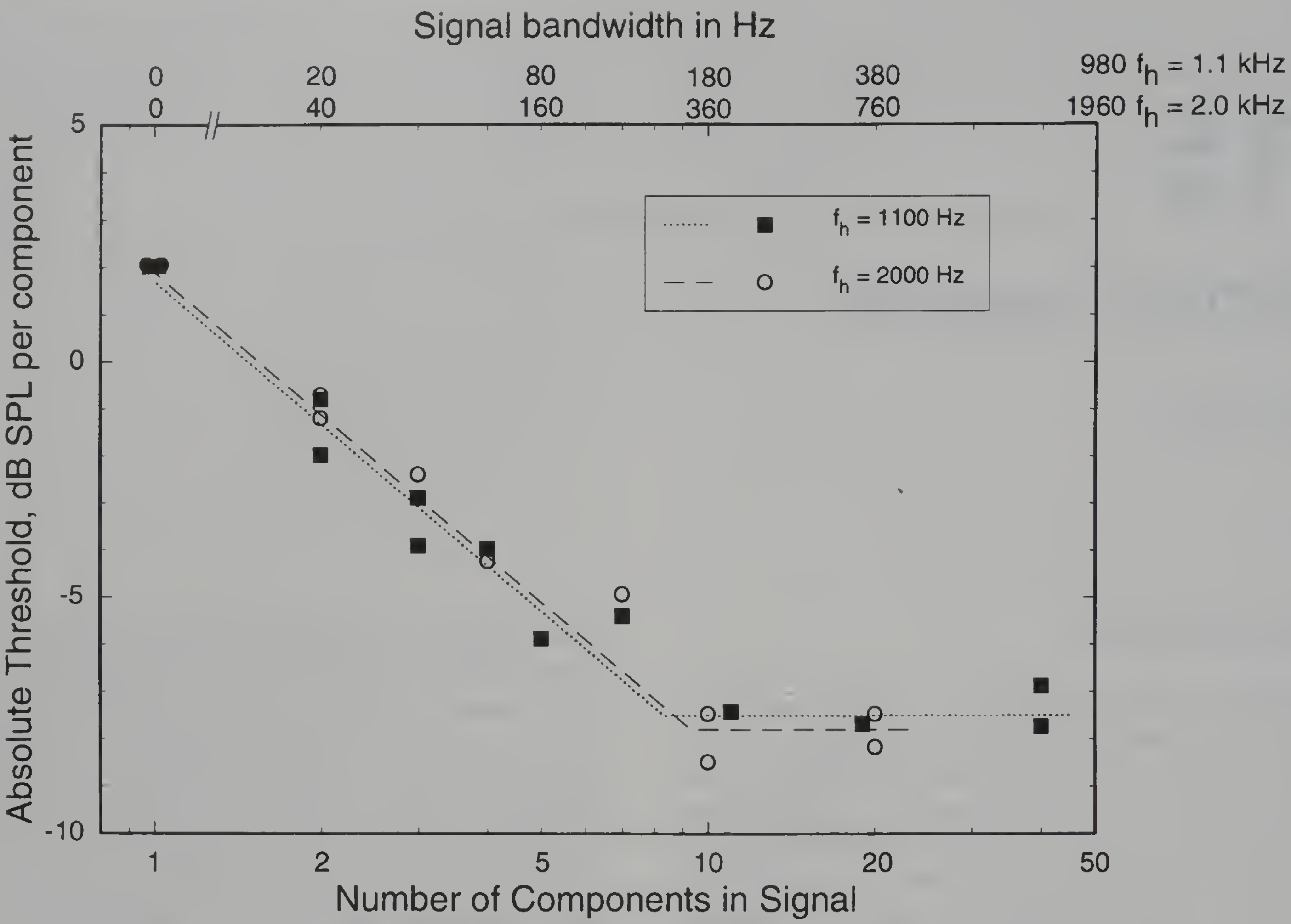


Fig. 1 Absolute thresholds as a function of the number of closely spaced, equally intense tones contained in the signal.[1] Data are shown for tones added every 20 Hz below an 1100-Hz upper frequency (■, dotted line) and for tones added every 40 Hz below a 2000-Hz upper frequency (○, dashed line). The corresponding bandwidths are indicated on the top abscissa. The lines are Gässler's summary of the data. The sloping line decreases 3 dB per doubling of the number of components, corresponding to perfect energy summation of signal components within the critical bandwidth. The horizontal lines indicate the thresholds that would be obtained if components more than 160 Hz below 1100 Hz and 320 Hz below 2000 Hz did not aid detection in the quiet. (Adapted with permission from Ref. 1.)

2.2 Effect of Duration

The threshold values discussed in Chapter 123 apply to signals with durations of 500 ms or more. Generally, thresholds increase for shorter durations.[3] For bursts of white noise with durations between 0.1 ms and 1 s, the intensity at threshold is well described by a power function of duration with an exponent of about -0.75.[4] This indicates that thresholds decrease about 7.5 dB for each 10-fold increase in duration. Tones with durations between the reciprocal of the auditory filter bandwidth and 500 ms yield similar results.[3] However, only part of the rise and fall may contribute to detection. Temporal integration data for multiple tone bursts follow a power function of duration more closely when only part of the rise and fall is included in the calculation of stimulus duration.[5]

Although power functions provide a good description of the effect of duration, they fail at very short and at long durations. Most studies show a faster rate of threshold decrease for durations less than about five periods of the tone frequency and a slower rate for durations longer than 200 ms.[3,6] For short durations, spectral splatter and auditory filtering can account for the faster rate.[6] The changes at extreme durations are evident in Fig. 2, which shows a summary of normal listeners' data for tones at or near 0.25, 1, 4, and 14 kHz obtained in several studies.

Whether the effect of duration changes with frequency is not clear. Figure 2 indicates that the frequency dependence of temporal integration, if any, is small. The best-fitting time constant for an exponential integrator model decreases with increasing signal frequency in some studies,[6] whereas little or no effect of frequency is evident in other studies.[3] The shallow slope at 14 kHz

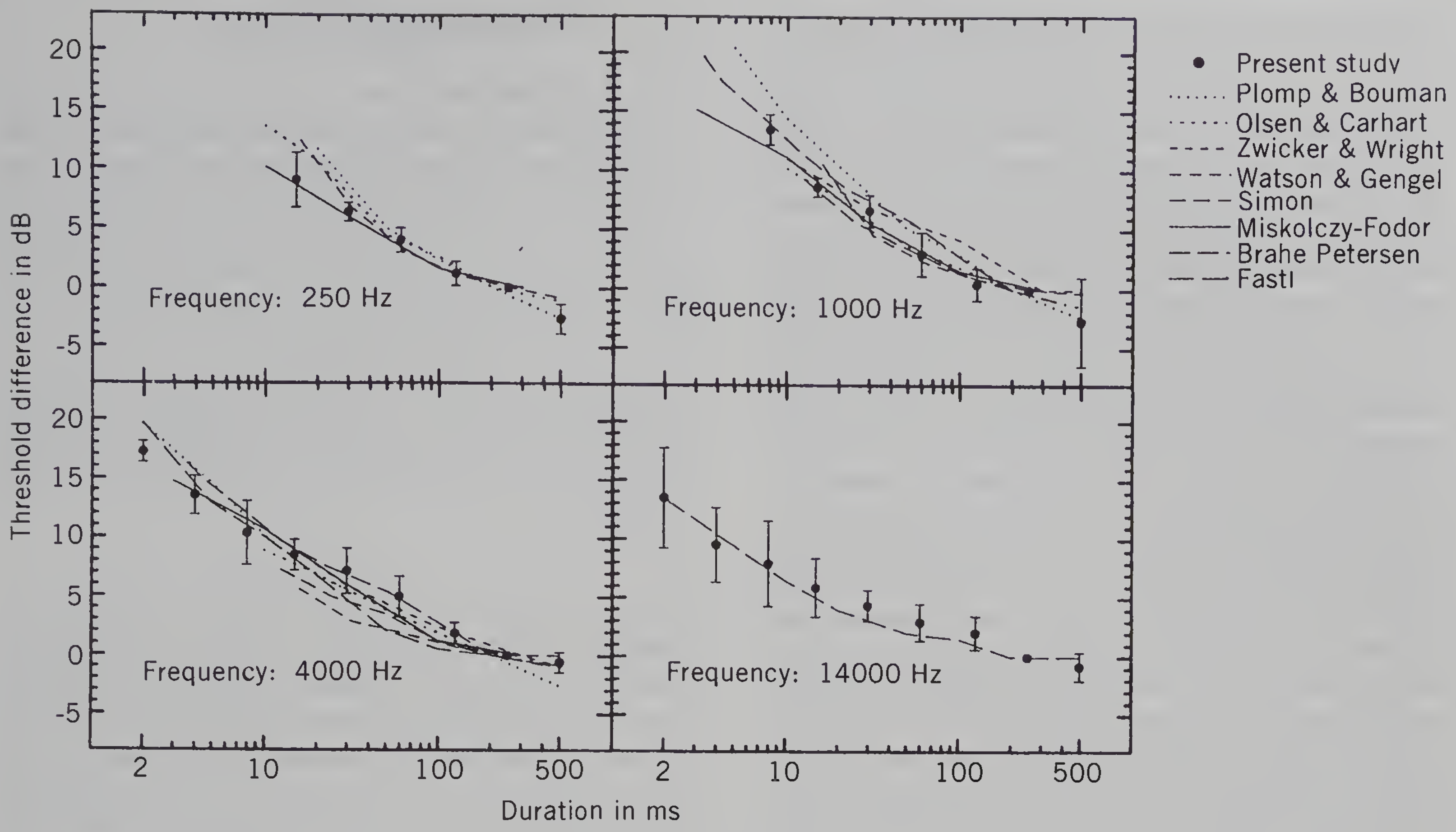


Fig. 2 Change in absolute thresholds as a function of duration for tones near 0.25 kHz (top left), 1 kHz (top right), 4 kHz (bottom left), and 14 kHz (bottom right) from Florentine et al.[3] The filled points show data for five normal listeners tested by Florentine et al.; the bars show plus and minus one standard deviation across listeners. The various lines show data obtained in other studies, as indicated on the right. (Reprinted with permission from Ref. 3.)

in Fig. 2 may be caused by elevated thresholds at this frequency in two of the five normal listeners tested.[3]

Data for detection of multiple brief tone bursts generally indicate that thresholds decrease as the number of bursts increases and increase as the presentation rate decreases, whether the overall signal duration or overall energy is constant.[7] However, one study found virtually no increase in thresholds as the interburst interval between four 8-ms bursts increased from 4 to 67 ms.[5] The reason for this discrepancy is unclear. Thresholds for two broadband clicks appear to increase from about −4 dB (relative to the threshold for a single click) for an interclick interval of 1 ms to an asymptote of approximately −1.5 dB for interclick intervals between 10 and 200 ms.[8]

The decrease in thresholds with increasing duration is often described as the result of a neural integration of stimulus power.[6] Plomp and Bouman's exponential integrator model predicts most of the data for multiple tone bursts and provides a good description of the effect of duration. In particular, it predicts that long-duration tones should show a gradual approach to the asymptotic threshold for continuous tones, which agrees with most data.[6]

An alternative account of the shallower slope at long durations comes from signal detection theory, which predicts that multiple independent observations of a long-duration tone should cause thresholds to decrease by 5 dB per decade of duration[9]—up to some limit set by decay of memory. The latter approach has the advantage that it easily accounts for findings that detection of a tone in noise improves up to a duration of 3 s.[9] The exponential integrator model cannot account for this finding unless the time constant is made considerably longer than the 100–300 ms normally used to fit the data.

However, models that integrate the stimulus intensity, such as the exponential integrator model, are inconsistent with the relatively small contribution to detection by the rise and fall. Moreover, this model cannot predict impaired listeners' data for either single or multiple tone bursts, although predictions obtained with a 10–30-ms time constant rarely differ more than 1 or 2 dB from the data. In part, this failure may reflect that impaired listeners can improve their performance by multiple observations on a stimulus longer than the integration time, but the threshold decrease is diminished due to these listeners' abnormally steep psychometric functions.[7]

Integration of stimulus intensity over long durations is also inconsistent with findings that the effect of interclick interval is asymptotic at around 10 ms.[8] This finding indicates that intensity is integrated only over a period less than 10 ms and that the decrease in threshold for longer durations reflects the effect of multiple observations. Such a model is consistent with the two-click data and with the finding that the threshold for two 10-ms tone bursts presented in temporal gaps of a broadband noise is largely unaffected by 6-dB changes in the level of the noise during 50 ms in the middle of a 100-ms interval that separates the gaps.[8] It also can predict the effect of duration for a single tone burst, but only if it is assumed that the weight on the signal increases with duration.

A somewhat different model entails short-term integration of a compressive power function transformation of the stimulus intensity.[4] The short integration time is compatible with the good temporal resolution shown in a variety of experiments (see Refs. 10 and 11 for reviews) and obviates the need to postulate different integration mechanisms for temporal integration and temporal resolution experiments. This model provides a good description of the effects of stimulus duration and interburst interval, but the parameters necessary to predict the latter effect differ from those used to predict the former. The model also predicts accurately the small effect of interburst interval obtained for impaired listeners[7] when its parameters are altered to reflect these listeners' abnormally fast growth of loudness and diminished temporal integration for single bursts. However, this model is inconsistent with the relatively small contribution to detection by the rise and fall.

None of the models specify the locus of temporal integration, which is most likely to be within or beyond the inferior colliculus because temporal integration is similar for acoustical stimulation and for electrical stimulation of the inferior colliculus.[12] Surprisingly, however, noise-induced hearing loss, which is thought to affect primarily the transduction from acoustical stimulus to neural impulses within the cochlea, changes temporal integration for both electrical and acoustical stimulation.[12] How lesions to the peripheral auditory system can result in changes to the apparently central process of integration is a mystery.

3 SIMULTANEOUS MASKING

A masker that is present throughout the duration of the signal produces simultaneous masking if it is sufficiently intense and its spectral energy is not too far removed from the signal. On-frequency masking occurs when the masker contains significant power in the frequency region of the signal.

3.1 On-Frequency-Noise Maskers

Effect of Masker Level Experiments on masking by white noise provide important information about the effects of masker level and signal frequency. The noise power density of white noise is independent of frequency. Thus, it has significant power in any frequency region—at least if the overall level is sufficiently high—and yields a simple relation between masker level and masked threshold. Comprehensive measurements show that the masked threshold varies with frequency and increases 10 dB for every 10-dB increase in masker level.[13] This linear relation is also obtained for narrowband maskers whose passbands encompass the signal frequency.[14–16]

Effect of Signal Frequency (Critical Ratios and Critical Bands) The masked thresholds of tones in white noise vary with frequency because on-frequency masking is determined primarily by the noise power within a narrow band of frequencies surrounding the signal and the width of that masking band varies with frequency, as discussed below. The idea that masked threshold is determined by the power at the outputs of the auditory filters is often referred to as the power spectrum model of masking. An elaborate version of such a model is discussed in Section 3.4.

Fletcher suggested that the effective signal-to-noise ratio within the masking band is equal to 0 dB at masked threshold.[14] According to this suggestion, the ratio of the spectral power density of the noise to the tone's intensity at masked threshold provides a direct measure of the range of frequencies that contribute to masking. Fletcher referred to this frequency range as a critical band.[14] However, other authors have reserved the term *critical band* for more direct estimates obtained in a wide variety of psychoacoustical experiments that vary the bandwidth of the stimulus (for review, see Ref. 17) and use the term *critical ratio* for the bandwidths measured by the ratio of the spectral power density of a white masking noise to the tone's intensity at masked threshold.

The critical band provides a basis for an auditory frequency scale—the Bark scale (named in honor of the German physicist Heinrich Barkhausen, who was a pioneer in the study of loudness). One bark is equal to the width of one critical band. Analytical expressions for the critical bandwidth as a function of physical frequency, $W_z(f)$, auditory frequency in barks as a function of physical frequency, $z(f)$, and physical frequency as a function of auditory frequency, $f(z)$, facilitate model calculations and planning of experiments. One such set of internally consistent expressions is given below. The critical bandwidth is

$$W_z(f) = \frac{(f+a)^b}{c}, \tag{1}$$

where $a = 1750$ Hz, $b = 1.81$, and $c = 9800\ \mathrm{Hz}^{0.81}$. (The unit for c is $\mathrm{Hz}^{0.81}$ to make the unit for $W_z(f)$ hertz.) The expression for $z(f)$ is obtained by integrating the reciprocal of Eq. (1):

$$z(f) = \frac{c}{1-b}\,\frac{1}{(f+a)^{b-1}}\ \text{barks} + k, \tag{2}$$

where the integration constant k is set 28.32 barks, such that z is zero when $f \approx 20$ Hz. Solving $z(f)$ for f yields

$$f(z) = \left(\frac{c \cdot 1\ \text{bark}}{(1-b)\cdot(z-k)}\right)^{1/(b-1)} + a. \tag{3}$$

Equation (1) yields values for $W_z(f)$ that are within ±10% of Zwicker's[18] table and Eq. (2) yields values for $z(f)$ that are within ±0.2 barks. Other expressions show similar deviations but are not internally consistent because they were derived by fitting each function separately.[19]

Figure 3 shows how the critical bandwidth and the critical ratio vary with frequency. The critical bandwidths are from Eq. (1) and the table.[18] The data for the critical ratios (converted to bandwidth under the assumption of a 0-dB signal-to-noise ratio at threshold) are for a signal duration of 100 ms and $d' = 1$.[20] The solid line is predicted by the model in Section 3.4. The critical ratios shown here are roughly equal to half the critical bandwidth, which is about 25% larger (1 dB in terms of thresholds) than those reported by Hawkins and Stevens.[13] The model indicates that this difference may reflect the different signal durations and threshold criteria used in the two studies. The difference between critical bandwidths and critical ratios usually is thought to reflect that the signal-to-noise ratio at threshold typically is −3 to −4 dB, rather than 0 dB.[17] Finally, the dashed line shows the equivalent rectangular bandwidth (ERB), which is calculated as $0.11(f + 165\ \mathrm{Hz})$, where f is the frequency in hertz (cf. Ref. 21). The ERB describes modern measurements of the auditory filter bandwidths in notched-noise masking experiments (discussed in Chapter 38 of Volume 2). The function for the ERB shows less curvature than the critical-band and critical-ratio functions. In the midfrequency range, the ERB is about 20%

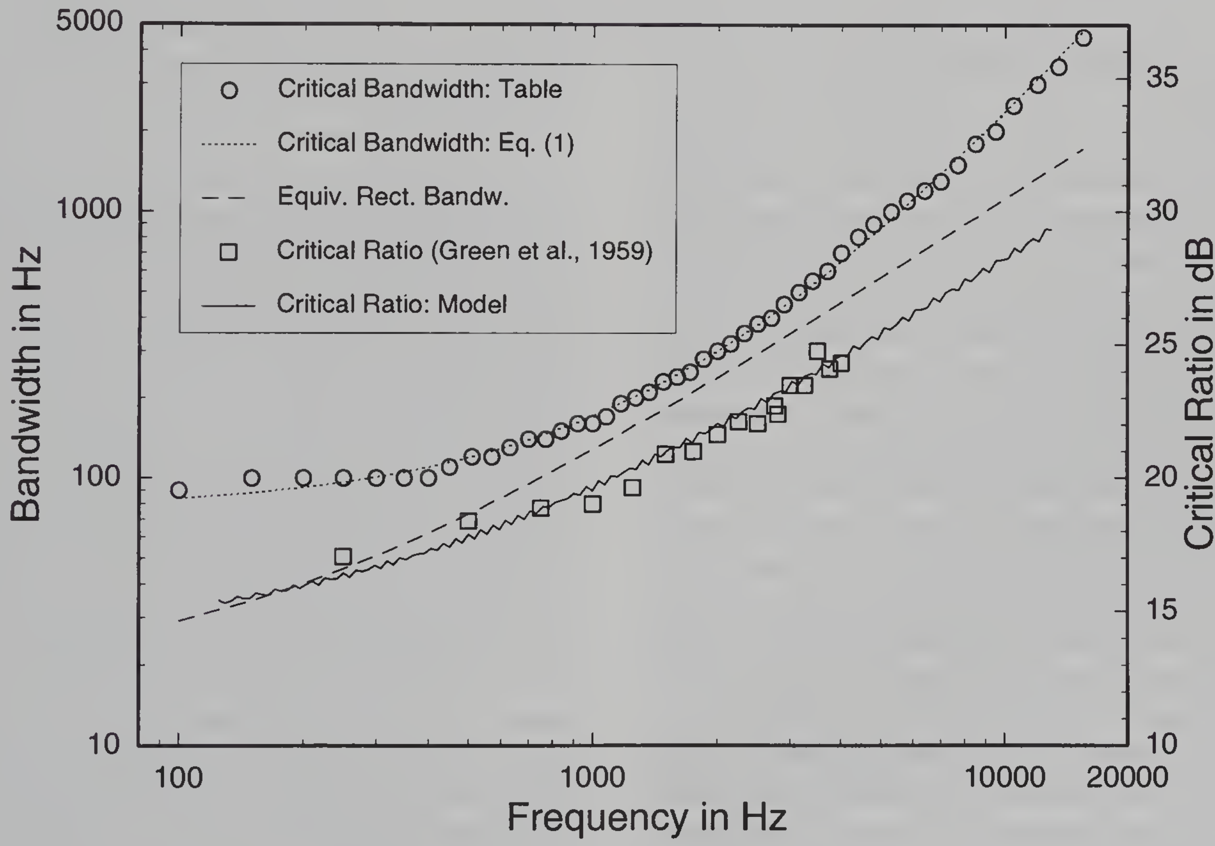


Fig. 3 Critical bandwidth, critical ratio, and ERB as a function of frequency. (○) Critical bandwidths from Zwicker[18] are compared to those calculated from Eq. (1) (dotted line) and to ERBs calculated according to a cochlear map function (dashed line).[21] (□) Critical ratios from Green et al.[20] are compared to predictions from the multiband masking model discussed in Section 3.4 (solid line).

less than the critical bandwidth; the difference increases at higher and lower frequencies.

Effect of Masker Bandwidth Fletcher envisioned the masking band as a bandpass filter centered on the signal frequency and stated that, as an approximation, it might be considered rectangular in shape.[14] If the signal-to-noise ratio at masked threshold is constant, this approximation predicts that thresholds in a variable-bandwidth masker with constant spectrum level should increase 3 dB for each doubling of bandwidth up to the bandwidth of the rectangular filter; further increases of masker bandwidth should have no effect. Several studies have obtained approximately such results.[14,22–24] However, the effect of masker bandwidth and the estimate of filter bandwidth vary considerably with the assumed shape of the filter. Moreover, signal detection theory indicates that the signal-to-noise ratio at masked threshold ought to increase as the masker bandwidth decreases because the relative variability of the noise energy in an observation increases.[22,25] The exact prediction depends on the relative magnitudes of variability in the auditory system and variability in the masker. Modern experiments show that thresholds for tones presented in narrow-band maskers with constant spectrum level decrease only about 1.5–2 dB per halving of masker bandwidth.[23,24]

Effect of Signal Bandwidth Like quiet thresholds, masked thresholds for complex signals also depend on their bandwidth. The masked threshold for a complex sound with closely spaced components is determined by the overall intensity as long as the bandwidth is less than that of the auditory filter.[1,2,26] For wider signal bandwidths, some studies indicate that detection is determined by the auditory filter with the best signal-to-noise ratio,[1,2] but most studies show that components in separate auditory filters are detected more readily than each component alone.[20,26–29] When the duration exceeds about 50 ms, the detectability of tone complexes with equally detectable components spaced more than a critical band apart increases as the square root of the number of components N. This $\sqrt{N}$ rule has been shown to hold for 2-tone complexes,[27] 12- and 16-tone complexes,[20] and 18-tone complexes.[28] Similar results have been obtained with noise bands as signals.[30]

The $\sqrt{N}$ (or square-root bandwidth) rule is easily understood by considering that the auditory system comprises many channels tuned to different frequencies. For tone complexes, it follows from a model using an optimal combination of observations in independent, frequency-selective channels.[27] For noise signals, the square-root bandwidth rule follows from signal detection theory because the relative stimulus variability decreases with increasing noise bandwidth. However, level discrimination of noise bands is not affected appreciably by stimulus variability unless it is quite large as when the bandwidth duration product (WT) is less than about 30.[31] Thus, it seems that the bandwidth effects obtained for noise signals are more likely to result from variations in internal noise and reflect the same multiband decision rules used to account for the $\sqrt{N}$ behavior obtained with complex-tone signals.

Serious departures from the $\sqrt{N}$ rule have been reported for signals with short durations, however. Complex signals composed of tone pulses with Gaussian envelopes and durations that vary with frequency to confine the energy of each component within a one-third octave yield almost complete integration of signal energy across three octaves.[29] The "efficient across-spectral integration" obtained for short-duration signals is difficult to explain by the multiband model.

3.2 Spread of Masking

If the auditory filters were rectangular, a band-limited masker should only mask signals whose frequencies are within its passband, but it is well known that intense low-frequency tones may mask high-frequency tones. This shows that the filters are not rectangular. The gradual attenuation on the skirts of the auditory filters is clearly shown by measurements of thresholds for tones at various frequencies in the presence of a fixed-frequency pure-tone masker.[32] As shown in Fig. 4, the results illustrate the fundamental properties of auditory filtering but are complicated by the many auditory percepts that arise from interactions of the masker and signal tones. The masked thresholds indicated by the solid line show relatively little masking at frequencies below the 1200-Hz masker but considerable masking above it. The notches around 2400 and 3600 Hz indicate that the ear creates harmonics by nonlinear distortion. When only a single sinusoid is presented, the harmonic distortion generally is not detectable, because it is masked by the stronger fundamental. However, adding a signal with frequency near a harmonic frequency produces beating, which reveals the existence of the aural harmonics. This beating causes the masked thresholds to show notches near the masker frequency and its second and third harmonics. Other aural distortion products—cubic-difference tones—enhance the detectability of signals between the masker frequency and its second harmonic.[33] The distortion products also affect perception well above the masked thresholds, and at higher signal levels, the percept of the masker plus signal becomes one of many components.

The asymmetry of masking indicated in Fig. 4 depends on the masker level. More masking is obtained

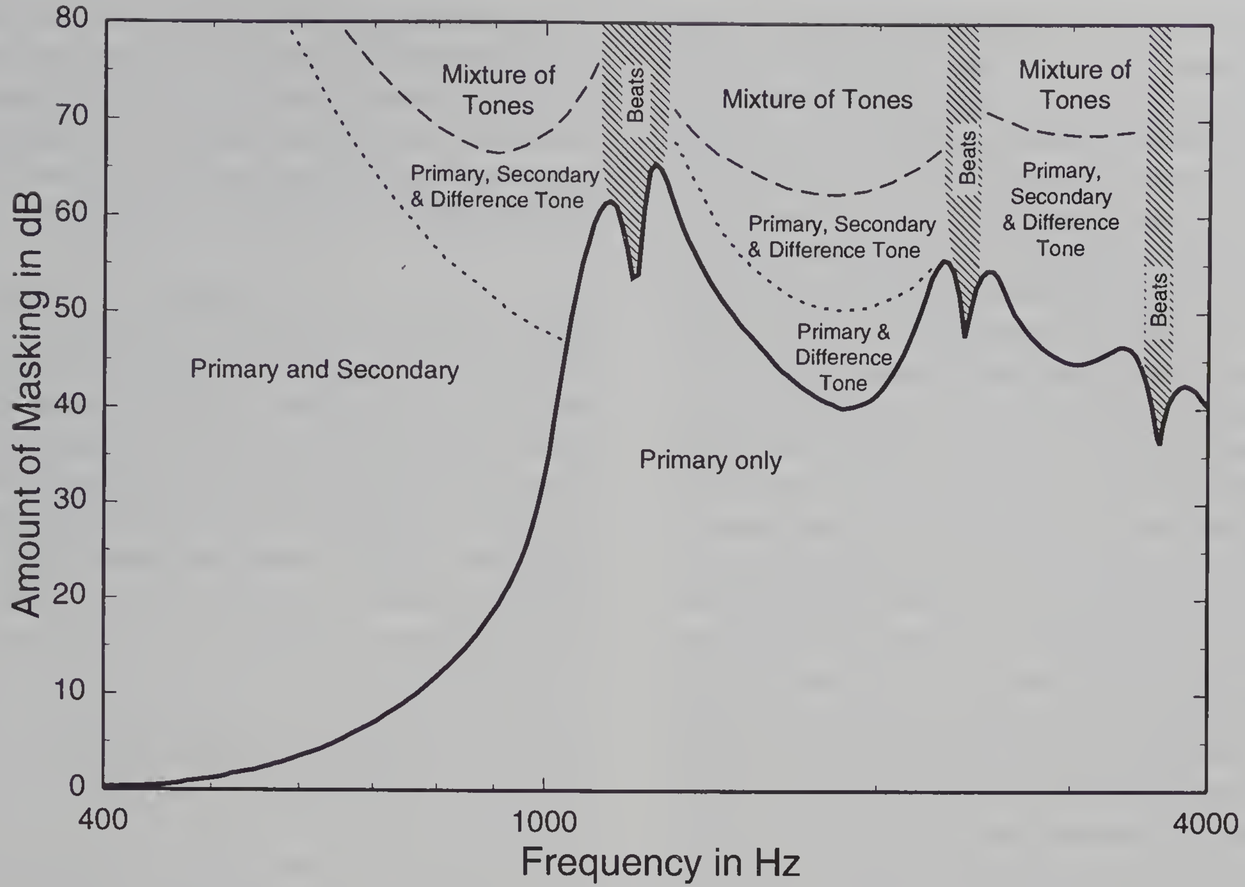


Fig. 4 Masked thresholds for tones at various frequencies in the presence of a 1200-Hz, 80-dB SPL tonal masker.[32] The solid line shows the masked thresholds as a function of frequency. Dashed and dotted lines show the approximate limits among the various percepts that result when the fixed masker is presented together with a signal of the frequency and level indicated. The hatched areas show regions of beating between the signal and the masker or masker harmonics created by aural distortion. (Reproduced with permission from Ref. 32.)

above than below a high-level masker, but the opposite holds for maskers below 40 dB SPL.[34] This reflects that masking by a pure-tone masker grows more slowly than the masker level for signal frequencies at or below the masker but faster than the masker level for signal frequencies above it.[15,16,32] For signal frequencies well above a pure-tone masker, masking may grow as much as 3 dB for each 1-dB increase in masker level. This faster-than-linear growth of masking probably reflects fundamental properties of basilar membrane mechanics. The amplitude of the basilar membrane vibration grows faster with amplitude of a tone when the frequency is below the characteristic frequency than when the frequency is near it (see Chapter 108).

Noise maskers with bandwidths less than a critical band generally yield masked thresholds quite similar to those obtained with a pure-tone masker with a level equal to the overall level of the noise,[15,16] but some differences are apparent. First, the random frequency and amplitude fluctuations of the noise masker render beats imperceptible. Second, the influence of aural distortion products is apparent for frequencies about one-half octave above the center frequency of a narrow-band masker, but it is considerably less than that obtained with pure-tone maskers. Third, aural distortion produces bands of noise below the masker by intermodulation distortion among the spectral components of the noise masker.[33] These bands elevate masked thresholds for signal frequencies somewhat below the masker's passband and probably are responsible for the "remote masking" that is observed far below the passband of intense octave-band maskers.[35] Finally, growth of masking for signal frequencies above that of the masker tends to be slower for noise maskers than for pure-tone maskers, and at high masker levels a pure-tone masker yields more masking of such signals than does a narrow-band noise.[15,36] The amplitude fluctuations of a narrow-band masker yield brief periods of low masker intensity, during which the listener may detect the tonal signal.[36] Certainly, thresholds for brief tones are lower when presented during periods of low masker intensity than during periods of high masker intensity of narrow-band maskers.[16]

In the experiments discussed so far, the level of the masker was kept constant and the signal varied to find the masked threshold. Other experiments fix the signal and vary the masker to find the level that just masks the signal for various masker frequencies.[37] When the signal is a pure tone at a low sensation level, the resulting plots of masker level as a function of masker frequency yield psychoacoustical tuning curves, which bear a striking resemblance to physiological tuning curves measured on the basilar membrane and in single auditory nerve fibers (see Chapters 108 and 109 for review of physiological tuning curves).[37] The low-level signal causes excitation only in auditory channels tuned to frequencies close to the signal and, insofar as masking is excitatory (i.e., is caused by activity due to the masker swamping the activity due to the signal), the psychoacoustical tuning curve might be considered an isoexcitation curve for the channels excited by the signal.

This simple view of psychoacoustical tuning curves is not tenable, however. Listeners can improve their detection performance by listening to frequencies other than that of the signal, for example, by detecting spectral splatter at the onset or offset of a tonal signal.[38] Even if the signal excites only a narrow range of auditory channels, listeners can improve detection of the signal by listening to channels tuned slightly above the signal frequency when the masker is below the signal and tuned slightly below it when the masker is above the signal, that is, "off-frequency listening."[39] Owing to the rounded shape of the auditory filter (see Chapter 116), the signal-to-masker ratio generally is best within a filter centered slightly away from the signal frequency. Furthermore, simultaneous masking may not be exclusively excitatory. Auditory nerve measurements indicate that it is excitatory for maskers above the signal frequency, but at least partly suppressive for intense maskers below the signal frequency.[40] Thus, it is likely that simultaneously masked tuning curves do not reflect an isoactivity contour for the masker's excitation in the channels tuned to the signal frequency. Rather, they may reflect an isosuppression contour for maskers below the signal frequency and an isoexcitation contour for maskers above it.

Although tone-on-tone masking and psychoacoustical tuning curves reveal basic properties of the auditory filters, they do not provide precise estimates of their shapes. A much more useful psychoacoustical characterization of auditory filtering is afforded by notched-noise maskers.[41] Briefly, the auditory filter shape is derived from masked thresholds for tones at frequencies between the proximal cutoff frequencies of two noise bands under the assumption that the thresholds represent a constant signal-to-noise ratio within the filter providing the best signal-to-noise ratio. This procedure takes the effect of off-frequency listening into account if detection is based on the linearly filtered signal energy, but not if detection is based on auditory distortion products arising from nonlinear interactions between the masker and the signal.[42] However, such nonlinear effects appear to be small in the conditions generally used to measure auditory filter shapes.[43] The properties of the auditory filters are discussed extensively in Chapter 116. As discussed in Section 3.4, rather precise estimates of masked thresholds for a variety of maskers and signals can be calculated on the basis of the outputs of such filters.

3.3 Additivity of Simultaneous Masking

The previous sections considered primarily maskers encompassing a single contiguous spectral region. What if two or more such maskers are combined? Except for effects of off-frequency listening, the power spectrum model of masking predicts that the total masking by the composite masker should be equal to the sum of the masking powers of the individual maskers. For example, consider two maskers that provide the same amount of masking. One might expect the amount of masking to increase 3 dB (i.e., the masking power to be doubled) when both maskers are presented together, but data show 10 or even 20 dB.[16,44,45] Whereas this *nonlinear additivity* or excess masking points to some limitations in the power spectrum model, a closer examination of the conditions used to produce it shows that it results from three factors: (a) off-frequency listening, (b) combination tone detection, and (c) changes in the masker envelope when two maskers are combined.[46]

The power spectrum model readily accommodates restrictions in off-frequency listening and can also account for the effects of distortion products if they are included as part of the effective stimulus. However, excess masking also occurs when off-frequency listening and auditory distortion products are ruled out, if at least one of the maskers fluctuates relatively slowly,[46] or if one masker is a low-frequency tone.[16] In these cases, it is likely that detectability of the signal in a single masker varies substantially over the course of a long-duration signal and that masked thresholds are based on brief periods of high detectabilty.[36,46] This effect cannot be explained by the power spectrum model, unless the weight assigned to the signal channel varies over time and depends on the short-term masker intensity.

3.4 Calculation of Simultaneous Masking

Although the traditional power spectrum model of masking fails in cases when the masker amplitude fluctuates slowly, it provides rather precise predictions of masked threshoslds in many cases. Therefore, it is worthwhile to consider a general version that can account for many of the data discussed above. The model is based on an energy detector model with internal noise.[25,30] It assumes that the stimulus is filtered by auditory filters with the characteristics discussed in Chapter 116 and that detection is based on the outputs of these filters, which are integrated over time and corrupted by internal noise. When the signal duration exceeds the integration time, multiple observations distributed over time are used to optimize performance. Likewise, multiple observations distributed across frequency-selective auditory channels are included to account for off-frequency listening and detection of complex signals.

To avoid having to select channels to suit a particular masker–signal combination, it is convenient to use channels that are spaced 0.5 ERBs apart. This is sufficient to keep the predicted thresholds for a fixed set of channels from deviating more than about 0.3 dB across different possible sets. Characterizing each channel by its center frequency $f_c(i)$, it can be shown that the overall sensitivity d'_{total} of such a multiband energy detector model is given as

$$d'_{\text{total}} = \left(\sum_{i=1}^{80} d'^2(f_c[i]) \right)^{1/2}, \tag{4}$$

where the center frequencies are

$$f_c(i) = (165\ \text{Hz}) \times (10^{0.024i} - 0.85) \tag{5}$$

and $d'(f_c)$ is the sensitivity to the signal within the auditory channel tuned to f_c, which can be calculated as

$$d'(f_c) = \begin{cases} \dfrac{T_s}{\tau} \dfrac{I_s(f_c)}{I_m(f_c)} \\ \quad \cdot \left(\dfrac{W(f_c)\tau}{\sigma_I^2 W(f_c)\tau + 2[1 + T_s I_s(f_c)/\tau I_m(f_c)]} \right)^{1/2} \\ \qquad \text{if } T_s \leq \tau, \\ \sqrt{\dfrac{T_s}{\tau}} \dfrac{I_s(f_c)}{I_m(f_c)} \\ \quad \cdot \left(\dfrac{W(f_c)\tau}{\sigma_I^2 W(f_c)\tau + 2[1 + I_s(f_c)/I_m(f_c)]} \right)^{1/2} \\ \qquad \text{if } t_s > \tau, \end{cases} \tag{6}$$

where T_s is the signal duration, τ is the 100-ms auditory integration time for the rectangular integration window in the model, $I_s(f_c)$ is the signal intensity and $I_m(f_c)$ the masker intensity at the output of the auditory filter centered on f_c, $W(f_c)$ is the ERB at f_c (see Chapter 116), and σ_I is a constant that determines the amount of internal noise. If σ_I is set to 0.65, this model yields predictions in excellent agreement with a wide variety of data. (The factor of 2 in front of the second term of the denominators accounts for the correlation between the overlapping channels.)

The model agrees with a large body of masking data, but this does not indicate that detection of a signal in noise necessarily is based on the stimulus energy. The critical ratio is nearly unaffected by random variations in the masker level between presentations, and the thresholds obtained with a 32-dB range are considerably lower than those predicted by an ideal energy detector.[47] Similar results have been obtained with narrow-band maskers.[48] These findings indicate that an across-interval comparison of energy in the signal band is an unlikely cue for detection of a tone in noise, contrary to the basic tenet of the power spectrum model of masking. Thus, these findings present a fundamental problem for the model and provide a strong basis for rejecting it. Nevertheless, the power spectrum model described by Eqs. (4)–(6) appears very useful as an engineering tool, and provides remarkably accurate predictions of listeners' performance in a wide variety of masking situations.

4 NONSIMULTANEOUS MASKING

The previous section discussed simultaneous masking as obtained when the masker is essentially continuous. However, a masking noise of limited duration also masks brief signals presented before the onset of the masker (backward masking) and, especially, following the offset of the masker (forward masking). Nonsimultaneous masking depends strongly on the temporal relations between the masker and the signal, as well as on the parameters that govern simultaneous masking.

4.1 Forward Masking

Temporal Parameters Forward masking decays as the delay between the masker and the signal offsets is increased, and little masking occurs beyond 200 ms.[49–51] The masker–signal delay is specified between offsets, because many investigators argue that this is the relevant variable.[49,52] The amount of masking is nearly independent of signal duration when the offset-to-offset delay Δt is fixed.[49] Thus, temporal effects in forward masking are most easily described by the amount of masking as a function of Δt.

The rate of decay in forward masking increases with the amount of masking produced for short delays. In other words, masked thresholds decrease faster with increasing masker–signal delay as the masker level and

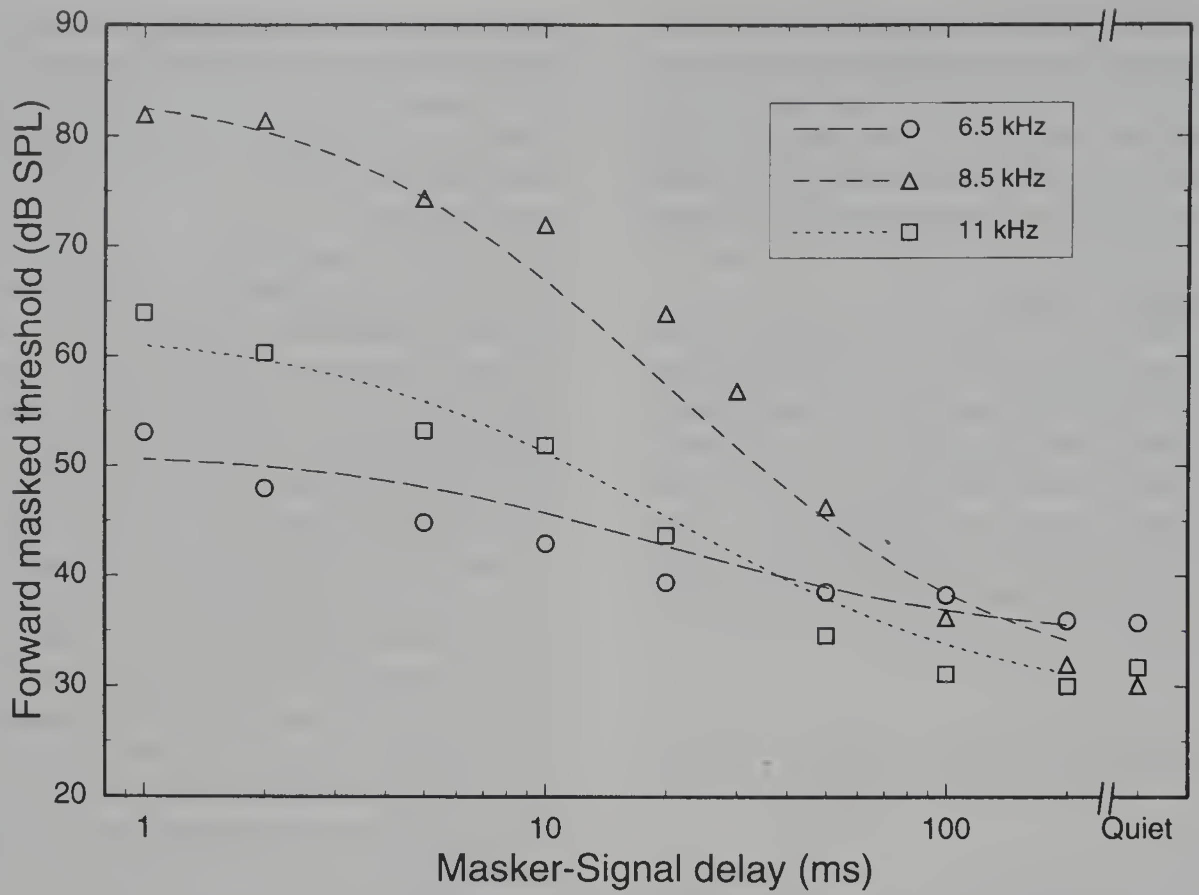


Fig. 5 Forward masked thresholds for tones at (○) 6.5, (△) 8.5, and (□) 11 kHz as a function of the delay between the offset of the 500-ms narrow-band masker at 8.5 kHz and the offset of the 1-ms signal.[50] Quiet thresholds are shown by the points to the far right. The dashed lines indicate masked thresholds calculated according to Eq. (7). (Adapted with permission from Ref. 50.)

spectral proximity of the masker and signal increase. This is illustrated in Fig. 5, which shows that forward masking by a long-duration masker lasts approximately 200 ms regardless of the initial amount of masking.[50] Generally, the decay of forward masking is well approximated by a straight line when the masked threshold in decibels is plotted as a function of delay on a logarithmic scale.[51,53] except for very brief delays, where masking shows almost no decay.[16,49,50] The rate of decay increases as the masker duration decreases below about 300 ms,[49,54,55] and forward masking ceases for signal delays exceeding about 50 ms when the masker duration is 5 ms.[55]

Because masking decays faster for high than for low masker levels, the effect of masker level depends on the masker–signal delay. For pure-tone maskers and signals with the same frequency, the interaction of masker level and temporal parameters in a forward-masking experiment can be expressed as a product of two independent factors.[51] The amount of masking increases proportionally with the masker level, and the proportionality constant is determined only by the temporal parameters. For example, the amount of masking at 1000 Hz increased about 0.6 dB for each 1-dB increase in masker level when Δt was 25 ms and about 0.25 dB when it was 60 ms.[51] These growth-of-masking slopes also depend on masker duration.[54]

The proportionality between masker level and amount of masking and the effects of temporal parameters may be summarized by a descriptive formula. Such a formula must reflect that forward masking degenerates into simultaneous masking as Δt goes to zero and into no masking as Δt becomes large; that is, the rate-of-growth multiplier describing the temporal effects in forward masking should decrease from unity (independent of masker duration T_m) for $\Delta t = 0$ to zero for Δt greater than about 200 ms. It should also reflect that the amount of masking obtained for a given Δt does not increase with T_m above 200–400 ms[49,50,54] and account for the effect of auditory filtering on forward masking.[11,56] One expression for the amount of masking M that fulfills these requirements is

$$M = \frac{d T_m^a}{d T_m^a + b(d + T_m)^a \dfrac{\Delta t^2}{\Delta t + c/\mathrm{ERB}(f_t)}} L_m^*, \qquad (7)$$

where a, b, c, and d are constants (d = 390 ms, a = 0.431, b = 0.495, and c = 0.693), ERB(f_t) is the auditory filter bandwidth at the test frequency f_t and L_m^* is the amount of simultaneous masking produced by the masker, which may be calculated from Eqs. (4)–(6). (The values of a, b, c, and d were obtained from a set 179 data points compiled from five studies.[16,49–51,54])

As shown by the lines in Figs. 5 and 6, this formula provides a very reasonable account of a variety of data on forward masking. The calculated amounts of masking rarely differ more than 3 or 4 dB from those measured. In particular, Eq. (7) describes quite well the interaction of masker duration and masker–signal delay shown in Fig. 6.[55] Both the data and Eq. (7) indicate that forward masking should be less than 3 dB when Δt exceeds 200 ms and the masker duration is 300–500 ms, and when Δt exceeds 50 ms and the masker duration is 5 ms.

Equation (7) also yields growth-of-masking slopes close to those observed in several studies.[50,51,54,57] Thus, Eq. (7) appears useful to estimate the amount of masking obtained in a variety of experiments, although its theoretical significance is unclear.

Frequency Selectivity Measured in Simultaneous and Forward Masking Frequency selectivity as measured by tuning curves, masking patterns, and auditory filter shapes generally is more acute in forward than in simultaneous masking. For example, forward-masked tuning curves have Q_{10} values (the signal frequency divided by the bandwidth of the tuning curve measured 10 dB above its tip) about twice those obtained in simultaneous masking and two to three times steeper slopes.[50,58] Likewise, forward-masked auditory filter characteristics have slopes almost twice those obtained under simultaneous masking, but the ERBs are only about 20% smaller.[41] In contrast, masking patterns are broader in forward than in simultaneous masking owing to a flattening of the peak that results because forward masking decays faster when the amount of simultaneous masking is large than when it is small.[50]

In part, the difference between forward and simultaneous masking may reflect the operation of suppression as has been observed in physiological measurements. In the auditory nerve, the response to a probe tone can be reduced or eliminated by introducing a second tone (the suppressor) at a frequency above or below that of the probe (see Chapter 108). Thus, a simultaneous masker may suppress the response to the signal. A forward masker cannot suppress the response to the signal because suppression appears to reflect instantaneous nonlinear interaction at the level of the basilar

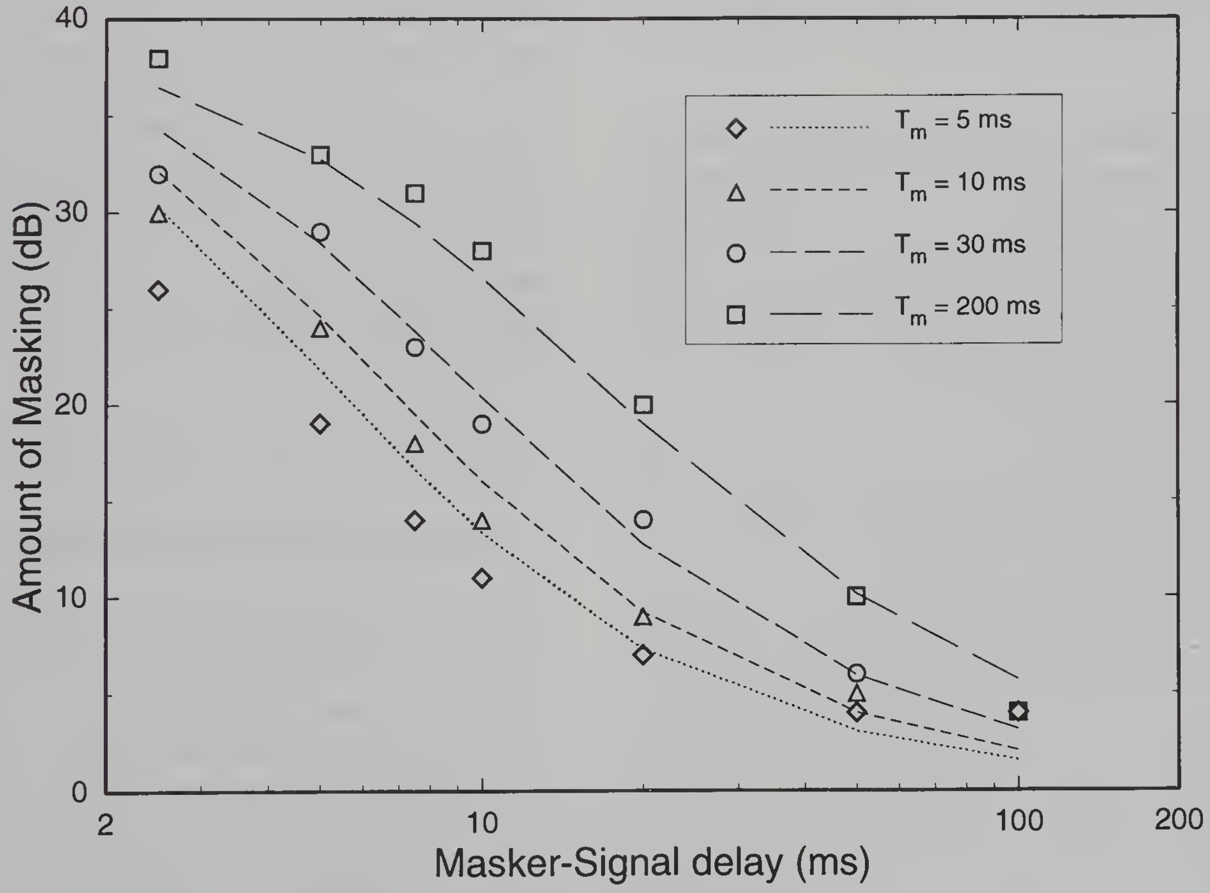


Fig. 6 Forward masking of 5-ms, 2-kHz tones preceded by bursts of uniformly masking noise are plotted as a function of the delay between masker and signal offsets. The parameter is masker duration as indicated. The symbols are data from Zwicker.[55] The lines indicate masked thresholds calculated according to Eq. (7). (Adapted with permission from Ref. 55.)

membrane (see Chapter 108). Accordingly, the forward-masked tuning curve may correspond to an isoexcitation pattern more closely than the simultaneously masked tuning curve.

The listerner's task in forward masking is quite complex, however, and a large part of the differences between forward and simultaneous masking is likely to reflect off-frequency listening and cueing.[59] In particular, listeners may have great difficulty in detecting a brief signal following an on-frequency forward masker. Excessive masking can occur if the delay between the masker and signal is too short to make the silence between them audible or if the delay and signal duration combine to make the signal confusible with natural fluctuations in a narrow-band masker.[60] Cueing and off-frequency listening are highly likely to affect forward-masked psychoacoustical tuning curves but are unlikely to affect the auditory filter measurements in forward masking, which employ maskers that have components both above and below the signal frequency and bandwidths much wider than that of the auditory filter centered on the signal. Consequently, the difference between forward and simultaneous masked auditory filter shapes probably is primarily due to suppression. This conclusion is consistent with the finding that the difference between forward and simultaneous masking is pronounced on the skirts of the auditory filters but modest in the central part of the filter characteristic, which provides the major contribution to the ERB.

Composite Maskers: Multiple Spectral Components Why is suppression not observed in simultaneous masking? One explanation is that, in simultaneous masking, the suppressor reduces activity due to the signal in the same proportion as that due to the masker.[61] Thus, suppression does not alter the effective signal-to-masker ratio in simultaneous masking and the masked threshold is not altered. In forward masking, the suppressor is presented together with the masker only, and thus, it does not suppress the signal. Consequently, suppression of the activity due to the masker improves the detectability of the signal. For example, forward masking by an on-frequency masker is reduced if a relatively intense suppressor is added at frequencies around one octave below or about 15% above the on-frequency masker.[62] The area of frequencies and SPLs for which suppression is obtained is qualitatively similar to the suppression areas observed in auditory nerve recordings. Suppression also occurs for maskers with broad spectra. Thresholds for a 2-kHz signal can be 20 dB lower with a 4000-Hz-wide masker than with a 500-Hz-wide masker,[63] but a large part of the decrease is likely to be caused by cueing effects as discussed above. An extensive review of the effects of cueing on estimates of suppression leads to the conclusion that the true contribution of suppression toward forward-masked thresholds rarely exceeds 5 dB.[59]

4.2 Backward Masking

A masker elevates thresholds not only for signals presented simultaneously with or after it but also for signals presented before the masker. The time interval for such backward masking rarely exceeds 25 ms[49] but can reach 100 ms.[64] Backward masking is frequency selective[50,63,64] and shows suppression,[63] which may reflect temporal smearing of the masker and signal by the auditory filters.[56] However, it may also reflect confusion between the masker and signal, because highly practiced listeners often show little or no backward masking.[65]

4.3 Nonadditivity of Forward, Simultaneous, and Backward Masking

Unusual masking effects can be obtained when masker components are separated in time. Two equally effective maskers presented without temporal overlap can produce 10 dB more masking than either masker alone.[66] Essentially similar results are obtained for two forward maskers, two backward maskers, and combinations of forward and backward maskers. Excess masking also has been shown for combinations of forward and simultaneous maskers (for a review, see Ref. 67). The excess masking may result because temporal integration in the auditory system acts on a compressive transformation of the stimulus power.[66] Compression by a modified power law can account for the nonaddivity of masking obtained with combinations of maskers that do not overlap temporally and/or spectrally,[67] but the exponent of the compressive power function is a free parameter that varies among listeners and experiments and has no obvious relation to other auditory phenomena.

5 SPECIAL EFFECTS IN MASKING

5.1 Interaural (Central) Masking

The masking effects discussed previously all were obtained with the masker and the signal presented to the same ear. Masking also can be produced when the masker is in one ear and the signal in the other. This interaural or central masking typically is measured using special earphones with more than 80 dB interaural attention, which ensures that the masking is not caused by the sound crossing from the masking ear to the signal ear (for review and references, see Ref. 52). Interaural masking grows very slowly (<0.2 dB/dB) with increasing masker level and is highly dependent on the delay between the masker and signal onsets. It is large when they are simultaneous but decreases with a time constant of 50 ms as the signal onset is delayed. When the masker is continuous, interaural masking is nearly absent for a pulsed signal but large for a continuous signal. Frequency selectivity is quite sharp. With 12–15 dB on-frequency interaural masking, the main peak of the masking pattern is about one critical band wide and masking decreases to less than 5 dB for signal frequencies more than about 250 Hz away from the masker. Unlike ipsilateral masking, interaural masking patterns are quite symmetrical up to a masker level of 60 dB SL (sensation level) but some asymmetry is apparent for 70–dB-SL maskers.

5.2 Comodulation Masking Release

The model of simultaneous masking presumes that detection of the signal depends only on the signal-to-noise ratio in channels responding to the signal. Thus, masking should depend primarily on the noise power in channels tuned to frequencies at or near a pure-tone signal. Whereas this simple view of masking is accurate in most instances, it fails under special circumstances. For example, the amount of masking may decrease from 80 to 40 dB when a continuous one-octave masker at 90 dB SPL is pulsed (32 ms on, 32 ms off)[68] but the masker energy decreases by only 3 dB. Because the masker excitation in the signal channel is reduced dramatically during the off periods, this release of masking may be thought to reflect temporal resolution.[68] According to this view, however, the same release should be obtained when a narrow-band masker is pulsed or modulated, which is not true.[69,70] This indicates that the release from masking depends, at least in part, on modulation at frequencies remote from the signal.

The release of masking is called comodulation masking release (CMR) because it occurs primarily when the modulation at remote frequencies is similar to that around the signal frequency, that is, when the on frequency and remote masker bands are comodulated.[23] As shown in Fig. 7, the effect of masker bandwidth changes dramatically when a Gaussian noise masker is multiplied by a 50-Hz low-pass noise, which produces envelopes that are approximately the same within each 100-Hz-wide band of noise. The data for random noise follows the pattern established by Fletcher.[14] As the bandwidth is increased, masked thresholds increase until some critical bandwidth is reached and then remain constant. The modulated noise yields results similar to those for the random noise up to a bandwidth of about 100 Hz. As the modulated noise is widened further, however, thresholds

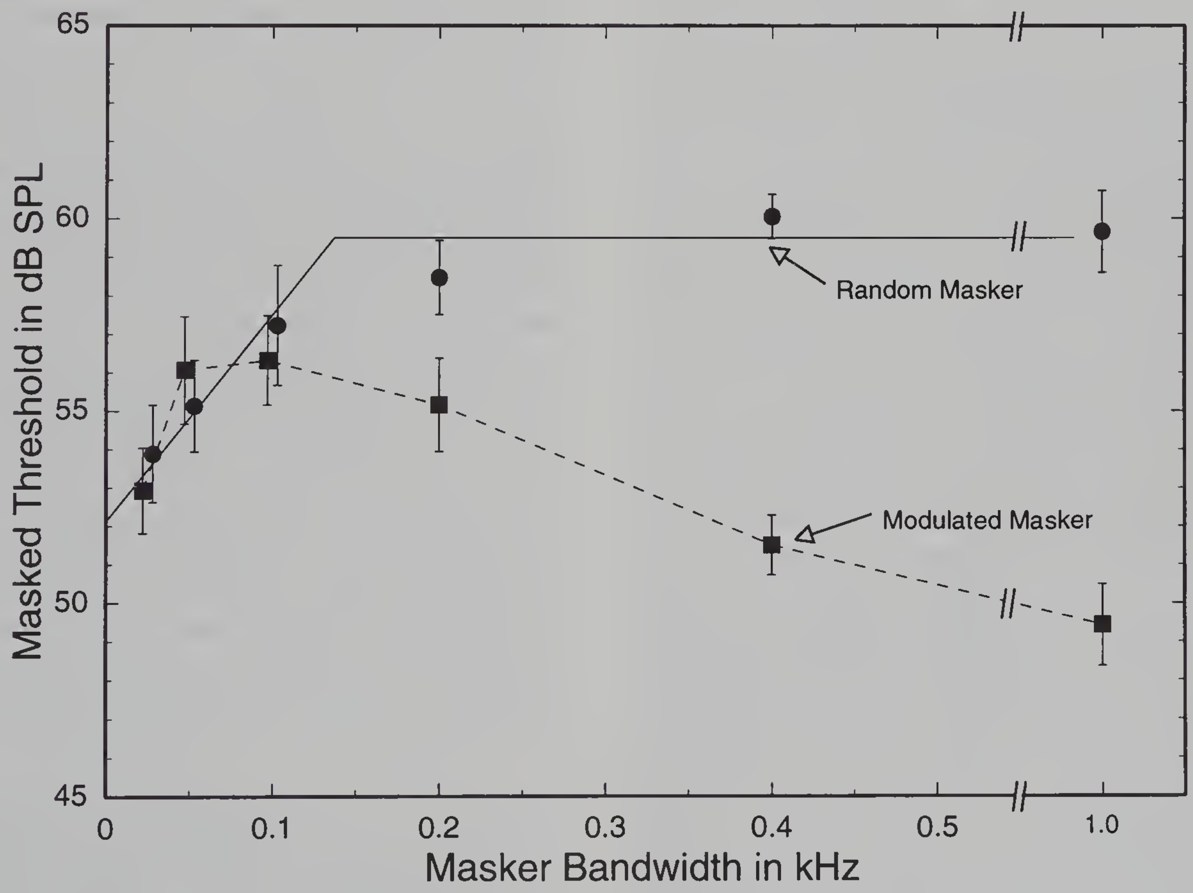


Fig. 7 Thresholds for 350-ms tones at 1 kHz masked by bands of random (●) or narrow-band noise-modulated (■ connected by dashed lines) noise plotted as a function of bandwidth of the masker.[23] The maskers were centered at 1 kHz. The bars show plus and minus one standard deviation across listeners. (Reproduced with permission from Ref. 23.)

start to decrease and at the 1-kHz bandwidth the common modulation results in a 10-dB reduction of masking.

The existence of CMR demonstrates that masking not only depends on the signal-to-noise ratio in channels excited by the signal but also may depend on channels excited only by the masker. When the masker consists of two or more bands with similar envelopes, masking is less than that obtained for a single narrow band centered on the signal. (The masker bands that are added away from the signal frequency are usually called cue bands.) The advantage afforded by CMR is most evident for detection. Level discrimination thresholds for partially masked tones of equal SPLs are approximately the same whether the masker is random or comodulated, except when the tones are at or below detection threshold in the random noise.[71] Likewise, filtered speech signals consistently yield a CMR for detection, but not for recognition.[72]

More than two dozen papers describe the effects of the many parameters in CMR experiments. Hall and Grose[70] provide an excellent review. Briefly, CMR generally decreases as the frequency separation between the on-frequency masker band and the cue band(s) increases.[73] In part, this effect may reflect that within-channel cues exist when a single cue band is separated from the on-frequency masker by less than one or two ERBs.[24] Several studies show that CMR increases as the bandwidth (or modulation frequency) of the masker band decreases.[74,75] In part, this apparent increase in CMR may be attributed to within-channel processes, because the contribution to CMR by the cue bands is almost independent of modulation frequency up to 100 Hz.[69]

If the cue bandwidth measured in ERBs or critical bands is held constant, CMR is approximately independent of frequency.[74] The level of the cue band relative to the masker band is of little importance to the magnitude of CMR. Comodulation masking release is present when the masker and cue bands differ in level by 30 dB[76] or even 60 dB,[77] but CMR increases with the masker level.[68,77] The range of masker–cue level differences for which CMR is obtained appears to be limited primarily by masking. As long as the cue does not make the masker inaudible or vice versa, a CMR is obtained. The wider the frequency separation between the masker and cue, the larger the range of level differences yielding CMR and the largest ranges are obtained when the cue band(s) are presented contralaterally to the masker and signal.[77]

Exact synchrony between masker and cue envelopes appears very important for CMR. If the envelope of the cue is phase shifted or delayed relative to the masker envelope, CMR is absent or reduced substantially.[78] Auditory grouping (i.e., the perception of disparate spectral components as a complex sound emanating from a single source) is important for CMR. Adding two codeviant bands, whose identical envelopes differ from that of six cue bands and the masker band, reduces the CMR, but it increases again as the number of codeviant bands increases.[79] Apparently, lack of perceptual segregation of a small number of codeviant bands causes them to be grouped with the cue envelopes, which effectively reduces the correlation between the cue envelope and the masker envelope.

One possible explanation for CMR is that detection may be mediated by a simultaneous comparison of envelopes in different critical bands.[23] This view is supported by the finding that CMR generally does not depend on uncertainty about the cue correlation, except if both the cue bands are uncorrelated in the non–signal interval and comodulated in the signal interval.[80] The envelope comparison may be described by an equalization–cancellation (EC) model, which extracts envelopes in different critical bands, adjusts their root-mean-square (rms) amplitudes to be equal (equalization), and subtracts envelopes in channels remote from the signal from the envelope in the signal channel (cancellation).[36] In the comodulated condition, the envelopes are similar when the signal is absent but dissimilar when it is present. Thus, the presence of the signal is indicated by a large remainder after cancellation.

Detection of decorrelation between the envelopes in the signal and the cue bands may also explain CMR.[73,81] The predictions by this model and the EC model may be identical.[82] When the signal is absent, comodulation causes the correlation to be high; when the signal is present, it alters the envelope in the signal band, which in turn decreases the correlation. The decorrelation caused by a signal close to threshold in a typical CMR experiment is sometimes comparable to the just-detectable decrease in correlation of two bands of noise,[81] but the effects of masker bandwidth are opposite in CMR and decorrelation detection.[83]

Comodulation masking release may also be mediated by detection of a difference in modulation depth between the masker-plus-signal and the cue bands.[76] The tonal signal reduces the modulation depth in the signal band, because it tends to "fill in" the periods of low masker amplitude. Indeed, a CMR can occur when the signal changes the depth but not the pattern of modulation.[84] In addition, CMRs obtained in sinusoidally modulated maskers are also consistent with discrimination thresholds for changes in modulation depth.[85]

A rather different explanation for CMR is that the comodulated cue indicates the optimal times to listen and allows listeners to "listen in the valleys."[36] If periods with low masking power ("valleys") are weighted more heavily than periods with high masking power, quite efficient detection may result. Like the other models, listening in the valleys requires extraction of the envelopes in different frequency bands, but unlike them, it does not depend on direct across-frequency comparisons of envelopes.

Many experiments have attempted to distinguish among the various explanations for CMR. Experiments with signals consisting of multiple tones show that a CMR still is obtained if signal components are present in all noise bands, but the presence of a cue band without signal components improves the CMR.[86,87] Simulations with specific implementations of the EC model, the correlation model, and the listening-in-the-valleys model indicate that such results are most consistent with the EC model.[86] However, different implementations of the general concepts embodied in the various models may yield different results.

Indeed, different conclusions result from other experiments. For example, a CMR is also obtained when the signal is identical to the on-frequency masker.[70,88] This finding is consistent with a comparison of envelope amplitudes between the masker–signal band and the cue band. Decorrelation detection and the EC model would not be effective because the signal does not alter the correlation between the signal and cue bands. Likewise, listening in the valleys should be no better than applying time-invariant weights. On the other hand, CMR is also obtained when the level of the cue band is varied randomly between presentations.[88] In this case, direct comparisons of envelope amplitudes are ineffective and CMR may be mediated by decorrelation detection or listening in the valleys.

The most surprising evidence, however, comes from experiments comparing CMR for signals presented at the peaks of the masker envelope and signals presented in the valleys.[70,87] No CMR is obtained for signals at the peaks, whereas a copious CMR is obtained for signals in the valleys. Because either signal causes a substantial change to the envelope, the EC model and the decorrelation model predict that a CMR should be obtained in both cases. Thus, these experiments clearly favor listening in the valleys as an explanation for CMR.

In summary, a CMR can be obtained even if one or the other of the proposed mechanisms are rendered ineffective by some experimental manipulation. The auditory system clearly can process simultaneous envelopes in different frequency bands in several different ways, and probably, detection is based on whichever process yields the best performance.[70]

5.3 "Overshoot" and the Effect of Off-Frequency "Fringes" on On-Frequency Masking

The section on simultaneous masking did not consider possible effects of the temporal relation between a pulsed masker and the signal. Indeed, the power spectrum model of masking does not easily predict such an effect. However, thresholds for brief tones presented in a wide-band masker can be elevated more than 10 dB if the signal onset occurs within a few milliseconds after the onset of a pulsed masker.[89] This overshoot of masking gradually diminishes as the delay between the masker and signal onsets increases up to about 300 ms.[89,90]

The overshoot produced by a broadband noise is small or absent at low levels, increases to reach about 10–15 dB at spectrum levels around 20 or 30 dB SPL, and either stays constant[91,92] or decreases somewhat at higher levels.[93–95] The signal frequency may explain part of the difference between studies. One study found that the amount of overshoot decreased at high levels for a 4-kHz signal, but not for a 1-kHz signal.[94] Likewise, another study found that the overshoot obtained with a wide-band masker decreased at high levels for a 6.5-kHz signal but not for a 2.5-kHz signal.[92] Individual differences may also play a role because the amount of overshoot obtained for a given masker–signal combination varies considerably among listeners.[93]

Overshoot is also obtained when the frequency of a tonal masker differs from the signal frequency.[96] In contrast, no overshoot is obtained when the masker is centered on the signal frequency and has a bandwidth similar to that of the signal, whether the bandwidth of the signal is narrow[89,96] or wide.[91]

As shown in Fig. 8, the overshoot decreases systematically as the bandwidth of a noise masker decreases.[89,90] The overshoot is absent for masker bandwidths of one and two critical bands but increases at wider bandwidths to reach about 12 dB for a white-noise masker (24 barks wide). Generally, overshoot is obtained only when the masker contains energy outside the frequency range encompassed by the auditory filters responding to the signal.[89,90,96–98] In particular, overshoot seems to depend on the presence of masker energy in auditory filters tuned to frequencies above that of the signal.[99] One exception is that overshoot also is obtained in narrow-band maskers at high frequencies and medium levels,[92] which indicates that the overshoot for such stimuli may, at least in part, be due to mechanisms different from those operating at other frequencies and/or levels.

The importance of off-frequency masker energy for the overshoot is very clearly demonstrated by studies that divided a broadband masker into several spectral regions in which the onset could occur at different times. Such measurements show that a broadband masker does not produce overshoot when the masker onset occurs close to the signal onset only in a narrow band around the signal frequency,[98] except at high frequencies and medium levels.[92] On the other hand, if the masker onset is close to the signal onset only in frequency regions outside a

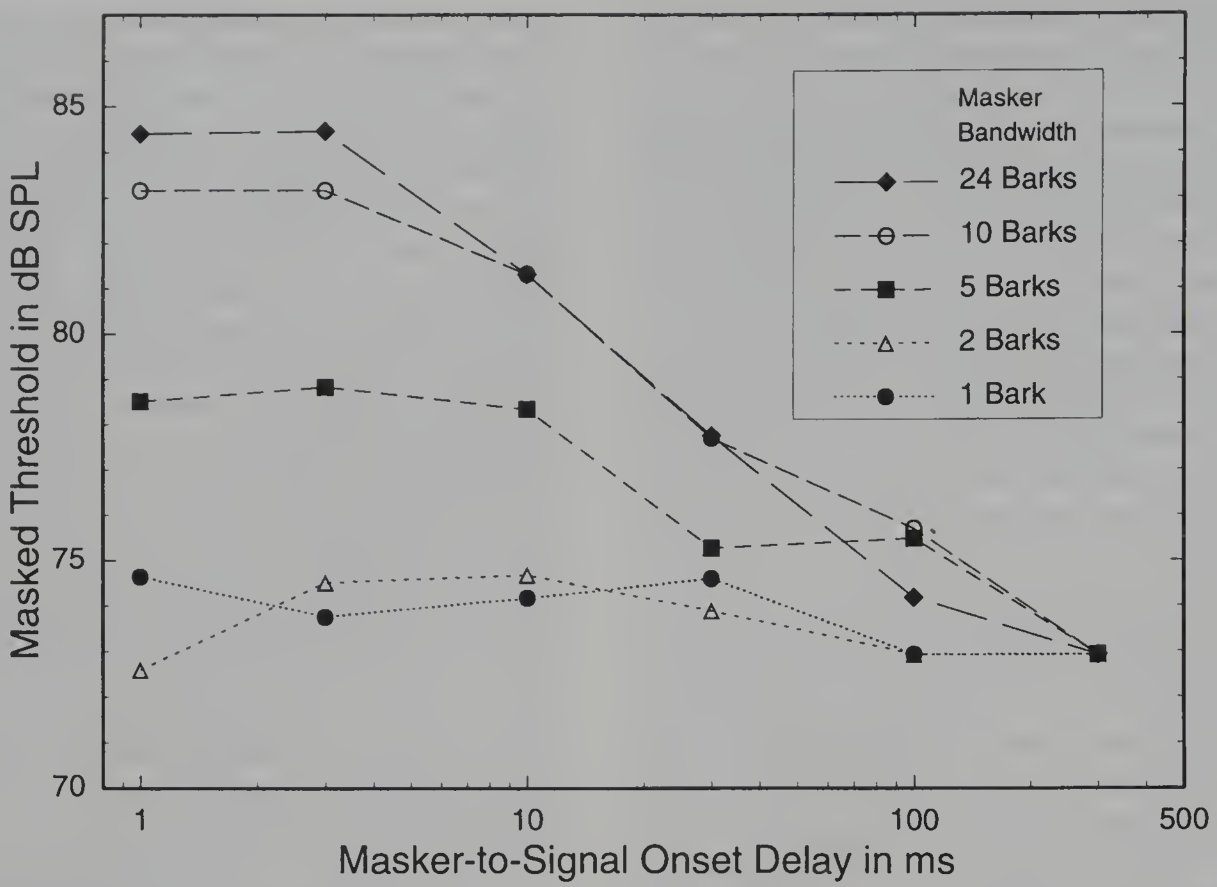


Fig. 8 Masked thresholds for a 2-ms tone at 4.8 kHz are plotted as a function of time between the onset of the 600-ms masker and the onset of the signal.[89] The parameter is masker bandwidth as indicated. (Reproduced with permission from Ref. 89.)

narrow band around the signal frequency, the overshoot is large.[92,98] Measurements with various cutoff frequencies between the on- and off-frequency bands indicate that the frequency regions important for producing overshoot with a 2.5-kHz signal are primarily between 1 and 2 kHz on the low side and between 3 and 4 kHz on the high side.[98] This corresponds to three to five ERB wide bands with proximal cutoff frequencies about one to two ERBs away from the signal frequency.

Probably several mechanisms are responsible for the overshoot. Several authors have suggested that overshoot is related to auditory nerve adaptation.[100] Although auditory nerve adaptation may contribute to the overshoot, it does not account for it completely because auditory nerve adaptation appears to be governed primarily by frequencies near the characteristic frequency,[101] whereas the overshoot is governed primarily by remote frequencies, at least at frequencies below 4 or 5 kHz.

Central mechanisms probably also are involved. Because one of the most important functions of the auditory system is to detect events in the environment as revealed by changes to the auditory input, it seems likely that any change in sound level and quality would attract attention. In terms of the power spectrum model, a sudden onset of the masker over a broad range of frequencies may temporarily force a high weighting to be applied to the channels remote from the signal. In essence, the sudden change in the masker channels distracts listeners from the task of monitoring the signal channel. As the onset in remote channels is removed in time from the signal onset, however, the change in the remote channels has been processed and attention can return to the signal channel. This hypothesis of time-varying channel weights is similar to the listening-in-the-valleys hypothesis for CMR and is consistent with findings that gating a 5-bark-wide masker causes a marked increase in the attention bandwidth as measured by the detectability of off-frequency probes.[102]

Auditory grouping may also contribute to the overshoot.[97] When the onsets of the masker and the signal occur together, their envelopes are quite similar and thus may be heard as a signal auditory object. This hypothesis is consistent with findings of slightly elevated thresholds for brief signals presented near the masker offset.[96] However, it may not be consistent with the overshoot's dependence on the off-frequency bands, because a signal should be about equally difficult to hear whether it is perceived as integral to a wide-band or a narrow-band auditory object.

Acknowledgments

Thanks are due to Mary Florentine, David M. Green, Armin Kohlrausch, Dennis McFadden, Fan-Geng Zeng, and two anonymous reviewers for helpful comments on an earlier draft. The preparation of this chapter was supported by NIDCD grant R01 DC00187.

REFERENCES

1. G. Gässler, "Über die Hörschwelle für Schallereignisse mit Verschieden Breitem Frequenzspektrum," *Acustica*, Vol. 4, 1954, pp. 408–414.
2. M. B. Higgins and C. W. Turner, "Summation Bandwidths at Threshold in Normal and Hearing-Impaired Listeners," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 2625–2630.
3. M. Florentine, H. Fastl, and S. Buus, "Temporal Integration in Normal Hearing, Cochlear Impairment, and Impairment Simulated by Masking," *J. Acoust. Soc. Am.*, Vol. 84, 1988, pp. 195–203.
4. M. J. Penner, "A Power-Law Transformation Resulting in a Class of Short-Term Integrators That Produce Time-Intensity Trades for Noise Bursts," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 195–201.
5. G. M. Gerken, V. K. H. Bhat, and M. Hutchison-Clutter, "Auditory Temporal Integration and the Power Function Model," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 767–778.
6. R. Plomp and M. A. Bouman, "Relation between Hearing Threshold and Duration for Tone Pulses," *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 749–758.
7. R. P. Carlyon, S. Buus, and M. Florentine, "Temporal Integration of Trains of Tone Pulses by Normal and by Cochlearly Impaired Listeners," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 260–268.
8. N. F. Viemeister and G. H. Wakefield, "Temporal Integration and Multiple Looks," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 858–865.
9. D. M. Green, T. B. Birdsall, and W. P. Tanner, "Signal Detection as a Function of Signal Intensity and Duration," *J. Acoust. Soc. Am.*, Vol. 29, 1957, pp. 523–531.
10. D. M. Green, "Temporal Factors in Psychoacoustics," in A. Michelsen (Ed.), *Time Resolution in Auditory Systems*, Springer, Berlin, 1985, pp. 122–140.
11. S. Buus and M. Florentine, "Gap Detection in Normal and Impaired Listeners: The Effect of Level and Frequency," in A. Michelsen (Ed.), *Time Resolution in Auditory Systems*, Springer, Berlin, 1985, pp. 159–179.
12. G. M. Gerken, J. M. Solecki, and F. A. Boettcher, "Temporal Integration of Electrical Stimulation of Auditory Nuclei in Normal-Hearing and Hearing-Impaired Cat," *Hear. Res.*, Vol. 53, 1991, pp. 101–112.
13. J. E. Hawkins and S. S. Stevens, "The Masking of Pure Tones and of Speech by White Noise," *J. Acoust. Soc. Am.*, Vol. 22, 1950, pp. 6–13.
14. H. Fletcher, "Auditory Patterns," *Rev. Mod. Phys.*, Vol. 12, 1940, pp. 47–65.
15. J. P. Egan and H. W. Hake, "On the Masking Pattern of a Simple Auditory Stimulus," *J. Acoust. Soc. Am.*, Vol. 22, 1950, pp. 622–630.
16. E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer, New York, 1990.
17. B. Scharf, "Critical Bands," in J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory*, Vol. 1, Academic, New York, 1970, pp. 157–202.
18. E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen)," *J. Acoust. Soc. Am.*, Vol. 33, 1961, p. 248.
19. E. Zwicker and E. Terhardt, "Analytical Expressions for Critical-Band Rate and Critical Bandwidth as a Function of Frequency," *J. Acoust. Soc. Am.*, Vol. 68, 1980, pp. 1523–1525.
20. D. M. Green, M. J. McKey, and J. C. R. Licklider, "Detection of a Pulsed Sinusoid as a Function of Frequency," *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 1446–1452.
21. D. D. Greenwood, "Critical Bandwidth and Consonance in Relation to Cochlear Frequency-Position Coordinates," *Hear. Res.*, Vol. 54, 1991, pp. 164–208.
22. E. De Boer, "Note on the Critical Bandwidth," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 985–986; Erratum, p. 1278.
23. J. W. Hall, M. P. Haggard, and M. A. Fernandes, "Detection in Noise by Spectrotemporal Analysis," *J. Acoust. Soc. Am.*, Vol. 76, 1984, pp. 50–56.
24. G. P. Schooneveldt and B. C. J. Moore, "Comodulation Masking Release (CMR) as a Function of Masker Bandwidth, Modulator Bandwidth, and Signal Duration," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 273–281.
25. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Academic, New York, 1966.
26. A. Langhans and A. Kohlrausch, "Spectral Integration of Broadband Signals in Diotic and Dichotic Masking Experiments," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 317–326.
27. D. M. Green, "Detection of Multiple Component Signals in Noise," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 904–911.
28. S. Buus, E. Schorer, M. Florentine, and E. Zwicker, "Decision Rules in Detection of Simple and Complex Tones," *J. Acoust. Soc. Am.*, Vol. 80, 1986, pp. 1646–1657.
29. W. A. C. Van den Brink and T. Houtgast, "Spectro-temporal Integration in Signal Detection," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 1703–1711.
30. C. E. Bos and E. De Boer, "Masking and Discrimination," *J. Acoust. Soc. Am.*, Vol. 39, 1966, pp. 708–715.
31. S. Buus, "Level Discrimination of Frozen and Random Noise," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 2643–2654.
32. R. L. Wegel and C. E. Lane, "The Auditory Masking of One Pure Tone by Another and Its Probable Relation to the Dynamics of the Inner Ear," *Phys. Rev.*, Vol. 23, 1924, pp. 266–285.
33. D. D. Greenwood, "Aural Combination Tones and Auditory Masking," *J. Acoust. Soc. Am.*, Vol. 50, 1971, pp. 502–543.

34. E. Zwicker and A. Jaroszewski, "Inverse Frequency Dependence of Simultaneous Tone-on-Tone Masking Patterns at Low Levels," *J. Acoust. Soc. Am.*, Vol. 71, 1982, pp. 1508–1512.
35. R. C. Bilger and I. J. Hirsh, "Masking of Tones by Bands of Noise," *J. Acoust. Soc. Am.*, Vol. 28, 1956, pp. 623–630.
36. S. Buus, "Release from Masking Caused by Envelope Fluctuations," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 1958–1965.
37. E. Zwicker, "On a Psychoacoustical Equivalent of Tuning Curves," in E. Zwicker and E. Terhardt (Eds.), *Facts and Models in Hearing*, Springer, Berlin, 1974.
38. B. Leshowitz and F. L. Wightman, "On-Frequency Masking with Continuous Sinusoids," *J. Acoust. Soc. Am.*, Vol. 49, 1971, pp. 1180–1190.
39. D. Johnson-Davies and R. D. Patterson, "Psychophysical Tuning Curves: Restricting the Listening Band to the Signal Region," *J. Acoust. Soc. Am.*, Vol.65, 1979, pp. 765–770.
40. B. Delgutte, "Physiological Mechanisms of Psychophysical Masking: Observations from Auditory-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 791–809.
41. B. C. J. Moore and B. R. Glasberg, "Auditory Filter Shapes Derived in Simultaneous and Forward Masking," *J. Acoust. Soc. Am.*, Vol. 69, 1981, pp. 1003–1014.
42. D. D. Greenwood, "Critical Band and Consonance: Their Operational Definitions in Relation to Cochlear Nonlinearity and Combination Tones," *Hear. Res.*, Vol. 54, 1991, pp. 209–246.
43. R. A. Lutfi and R. D. Patterson, "Combination Bands and the Measurement of the Auditory Filter," *J. Acoust. Soc. Am.*, Vol. 71, 1982, pp. 421–423.
44. D. M. Green, "Additivity of Masking," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 1517–1525.
45. R. A. Lutfi, "Additivity of Simultaneous Masking," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 262–267.
46. B. C. J. Moore, "Additivity of Simultaneous Masking, Revisited," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 488–494.
47. G. Kidd, Jr., C. R. Mason, M. A. Brantley, and G. A. Owen, "Roving-Level Tone-in-Noise Detection," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 1310–1317.
48. V. M. Richards, "The Detectability of a Tone Added to Narrow Bands of Equal-Energy Noise," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 3424–3435.
49. H. Fastl, "Temporal Masking Effects: I. Broad Band Noise Masker," *Acustica*, Vol. 35, 1976, pp. 287–302.
50. H. Fastl, "Temporal Masking Effects: II. Critical Band Noise Masker," *Acustica*, Vol. 36, 1977, pp. 317–331.
51. W. Jesteadt, S. P. Bacon, and J. R. Lehman, "Forward Masking as a Function of Frequency, Masker Level, and Signal Delay," *J. Acoust. Soc. Am.*, Vol. 71, 1982, pp. 950–962.
52. J. J. Zwislocki, "Masking: Experimental and Theoretical Aspects of Simultaneous, Forward, Backward, and Central Masking," in E. C. Carterette and M. P. Friedman (Eds.), *Handbook of Perception, Vol. 4: Hearing*, Academic, New York, 1978, pp. 283–336.
53. R. Plomp, "Rate of Decay of Auditory Sensation," *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 277–282.
54. G. Kidd, Jr., C. R. Mason, and L. L. Feth, "Temporal Integration and Forward Masking in Listeners Having Sensorineural Hearing Loss," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 937–944.
55. E. Zwicker, "Dependence of Post-masking on Masker Duration and Its Relation to Temporal Effects in Loudness," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 219–223.
56. H. Duifhuis, "Consequences of Peripheral Frequency Selectivity for Nonsimultaneous Masking," *J. Acoust. Soc. Am.*, Vol. 54, 1973, pp. 1471–1488.
57. B. C. J. Moore, P. W. F. Poon, S. P. Bacon, and B. R. Glasberg, "The Temporal Course of Masking and the Auditory Filter Shape," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 1873–1880.
58. B. C. J. Moore, "Psychophysical Tuning Curves Measured in Simultaneous and Forward Masking," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 524–532.
59. B. C. J. Moore and B. J. O'Loughlin, "The Use of Nonsimultaneous Masking to Measure Frequency Selectivity and Suppression," in B. C. J. Moore (Ed.), *Frequency Selectivity in Hearing*, Academic, New York, 1986, pp. 179–250.
60. D. L. Neff, "Confusion Effects with Sinusoidal and Narrow-Band Noise Forward Maskers," *J. Acoust. Soc. Am.*, Vol. 79, 1986, pp. 1519–1529.
61. T. Houtgast, "Auditory-Filter Characteristics Derived from Direct-Masking Data and Pulsation-Threshold Data with a Rippled-Noise Masker, *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 409–415.
62. R. V. Shannon, "Two-Tone Unmasking and Suppression in a Forward-Masking Situation," *J. Acoust. Soc. Am.*, Vol. 59, 1976, pp. 1460–1470.
63. D. L. Weber, "Suppression and Critical Bands in Band-Limiting Experiments," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 141–150.
64. L. L. Elliott, "Development of Auditory Narrow-Band Frequency Contours," *J. Acoust. Soc. Am.*, Vol. 42, 1967, pp. 143–153.
65. B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 3rd ed., Academic, London, 1989.
66. M. J. Penner and R. M. Shriffrin, "Nonlinearities in the Coding of Intensity within the Context of a Temporal Summation Model," *J. Acoust. Soc. Am.*, Vol. 67, 1980, pp. 617–627.
67. L. E. Humes and W. Jesteadt, "Models of the Additivity of Masking," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 1285–1294.
68. E. Zwicker, "A Device for Measuring Temporal Resolution of the Ear," *Audiol. Acoust.*, Vol. 19, 1980, pp. 94–108.
69. R. P. Carlyon, S. Buus, and M. Florentine, "Comodula-

tion Masking Release for Three Types of Modulator as a Function of Modulation Rate," *Hear. Res.*, Vol. 42, 1989, pp. 37–46.

70. J. W. Hall and J. H. Grose, "Relative Contributions of Envelope Maxima and Minima to Comodulation Masking Release," *Quart. J. Exp. Psych.*, Vol. 43A, 1991, pp. 349–372.
71. J. W. Hall and J. H. Grose, "Amplitude Discrimination in Masking Release Paradigms," *J. Acoust. Soc. Am.*, Vol. 98, 1995, pp. 847–852.
72. J. H. Grose and J. W. Hall, "Comodulation Masking Release for Speech Stimuli," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 1042–1050.
73. M. Cohen and E. D. Schubert, "Influence of Place Synchrony on Detection of a Sinusoid," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 452–458.
74. M. P. Haggard, J. W. Hall, and J. H. Grose, "Comodulation Masking Release as a Function of Bandwidth and Test Frequency," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 113–118.
75. D. A. Eddins and B. A. Wright, "Comodulation Masking Release for Single and Multiple Rates of Envelope Fluctuation," *J. Acoust. Soc. Am.*, Vol. 96, 1994, pp. 3432–3442.
76. J. W. Hall, "The Effect of Across-Frequency Differences in Masking Level on Spectrotemporal Pattern Analysis," *J. Acoust. Soc. Am.*, Vol. 79, 1986, pp. 781–787.
77. B. C. J. Moore and M. J. Shailer, "Comodulation Masking Release as a Function of Level," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 829–835.
78. B. C. J. Moore and G. P. Schooneveldt, "Comodulation Masking Release as a Function of Bandwidth and Time Delay between On-Frequency and Flanking-Band Maskers," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 725–731.
79. J. W. Hall and J. H. Grose, "Comodulation Masking Release and Auditory Grouping," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 119–125.
80. B. A. Wright and D. McFadden, "Uncertainty about the Correlation among Temporal Envelopes in Two Comodulation Tasks," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 1339–1350.
81. V. M. Richards, "Monaural Envelope Correlation Perceptions," *J. Acoust. Soc. Am.*, Vol. 82, 1987, pp. 1621–1630.
82. D. M. Green, "On the Similarity of Two Theories of Comodulation Masking Release," *J. Acoust. Soc. Am.*, Vol. 91, 1992, p. 1769.
83. B. C. J. Moore and D. S. Emmerich, "Monoaural Envelope Correlation Perception, Revisited: Effects of Bandwidth, Frequency Separation, Duration, and Relative Level of Noise Bands," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 2628–2633.
84. D. A. Fantini and B. C. J. Moore, "Comodulation Masking Release (CMR) and Profile Analysis: The Effect of Varying Modulation Depth," in Y. Cazals, L. Demany, and K. Horner (Eds.), *Auditory Physiology and Perception*, Pergamon, New York, 1992, pp. 479–488.
85. D. A. Fantini, "The Processing of Envelope Information in Comodulation Masking Release (CMR) and Envelope Discrimination," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 1876–1888.
86. J. W. Hall, J. H. Grose, and M. P. Haggard, "Comodulation Masking Release for Multicomponent Signals," *J. Acoust. Soc. Am.*, Vol. 83, 1988, pp. 677–686.
87. W. A. C. Van Den Brink, T. Houtgast, and G. F. Smoorenburg, "Signal Detection in Temporally Modulated and Spectrally Shaped Maskers," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 267–278.
88. J. W. Hall and J. H. Grose, "Comodulation Masking Release: Evidence for Multiple Cues," *J. Acoust. Soc. Am.*, Vol. 84, 1988, pp. 1669–1675.
89. E. Zwicker, "Temporal Effects in Simultaneous Masking and Loudness," *J. Acoust. Soc. Am.*, Vol. 38, 1965, pp. 132–141.
90. S. P. Bacon and M. A. Smith, "Spectral, Intensive, and Temporal Factors Influencing Overshoot," *Quart. J. Exp. Psych.*, Vol. 43A, 1991, pp. 373–400.
91. E. Zwicker, "Temporal Effects in Simultaneous Masking by White-Noise Bursts," *J. Acoust. Soc. Am.*, Vol. 37, 1965, pp. 653–663.
92. R. P. Carlyon and L. J. White, "Effect of Signal Frequency and Masker Level on the Frequency Regions Responsible for the Overshoot Effect," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 1034–1041.
93. S. P. Bacon, "Effect of Masker Level on Overshoot," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 698–702.
94. S. P. Bacon and G. A. Takahashi, "Overshoot in Normal-Hearing and Hearing-Impaired Subjects," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 2865–2871.
95. R. Von Klitzing and A. Kohlrausch, "Effect of Masker Level on Overshoot in Running- and Frozen-Noise Maskers," *J. Acoust. Soc. Am.*, Vol. 95, 1994, pp. 2192–2201.
96. S. P. Bacon and N. F. Viemeister, "The Temporal Course of Simultaneous Tone-on-Tone Masking," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 1231–1235.
97. R. P. Carlyon, "Changes in the Masked Thresholds of Brief Tones Produced by Prior Bursts of Noise," *Hear. Res.*, Vol. 41, 1989, pp. 223–236.
98. D. McFadden, "Spectral Differences in the Ability of Temporal Gaps to Reset the Mechanisms Underlying Overshoot," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 254–261.
99. S. Schmidt and E. Zwicker, "The Effect of Masker Spectral Asymmetry on Overshoot in Simultaneous Masking," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 1324–1330.
100. R. L. Smith and J. J. Zwislocki, "Short-Term Adaptation in Incremental Response of Single Auditory-Nerve Fibers," *Biol. Cybern.*, Vol. 17, 1975, pp. 169–182.
101. D. M. Harris and P. Dallos, "Forward Masking of Auditory Nerve Fiber Responses," *J. Neurophysiol.*, Vol. 42, 1979, pp. 1083–1107.
102. H. Dai and S. Buus, "Effect of Gating the Masker on Frequency-Selective Listening," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 1816–1818.

116

FREQUENCY ANALYSIS AND PITCH PERCEPTION

BRIAN C. J. MOORE

1 INTRODUCTION

Unlike the sinusoidal tones often used in the assessment of hearing, sounds encountered in everyday life are generally complex, containing many sinusoidal frequency components. One of the most important characteristics of the auditory system is that it acts as a limited-resolution frequency analyzer; complex sounds are broken down into their sinusoidal frequency components. The initial basis of this frequency analysis almost certainly depends upon the tuning observed in the cochlea (see Chapter 108). Indeed, it is possible that the tuning observed in the cochlea is sufficient to account for the frequency-analyzing capacity of the entire auditory system.[1] It is largely as a consequence of this frequency analysis that it is possible to hear one sound in the presence of another sound with a different frequency. This ability is known as frequency selectivity or frequency resolution.

This chapter starts by reviewing ways in which the frequency selectivity of the ear can be measured and quantified using masking experiments. It goes on to describe some other aspects of auditory perception that depend strongly on frequency analysis, specifically the ability to hear individual partials in complex sounds and the dependence of loudness on the bandwidth of sounds.

The second half of the chapter gives an account of the perception of pitch for pure tones and for complex tones. It argues that pitch perception depends partly on the frequency analysis that occurs in the peripheral auditory system. However, processes based on time pattern analysis also play an important role.

2 CHARACTERIZATION OF FREQUENCY-ANALYZING CAPACITY OF THE EAR

2.1 Masking and the Auditory Filter

Masking can be used to measure the limits of frequency analysis; when two sounds with different frequencies are completely resolved by the ear, then neither will be masked by the other. One conception of auditory masking, which has had both theoretical and practical success, assumes that the auditory system includes a bank of overlapping bandpass filters[2] (see Chapter 115). It is assumed that the observer will "listen" to the filter whose output has the highest signal-to-masker ratio. The signal will be detected if that ratio exceeds a certain value. In experiments where the bandwidth of the signal or masker is varied, the results often change at a certain bandwidth, called the *critical bandwidth*[3,4] (CB). This bandwidth corresponds approximately to the bandwidth of the auditory filter.

The relative response of the auditory filter as a function of frequency is often called the filter "shape." Patterson et al.,[5] based on the results of experiments measuring the detection of a sinusoidal signal presented in a noise masker with a spectral notch, suggested that the shape of each side of the auditory filter could be described approximately by the expression

$$W(g) = (1 + pg)\exp(-pg), \tag{1}$$

where g is the normalized frequency deviation from the center of the filter; $g = |f - f_c|/f_c$. The parameter p, calculated from the detection data, determines the sharpness of the auditory filter and its value may differ for the upper and lower halves. The equivalent rectangu-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

lar bandwidth (ERB) of the auditory filter is equal to $2f_c/p_l + 2f_c/p_u$, where p_l and p_u are the values of p for the lower and upper halves of the filter. At moderate sound levels p_l and p_u are approximately equal, but at high sound levels p_l tends to decrease; that is, the low-frequency side of the filter becomes less steep.[6]

Auditory filters have ERBs that increase with increasing center frequency. At moderate sound levels, the ERB is well approximated by

$$\text{ERB} = 24.7(4.37F_c + 1), \tag{2}$$

where F_c is frequency in kilohertz.[7] This expression is shown in Fig. 1, together with estimates of the ERB from several different experiments. (For quick mental calculations the reader might find the following approximate expression useful: $\text{ERB} = \frac{1}{9}f_c + 25$, where f_c is in hertz.) At a proportion of center frequency, the bandwidth tends to be narrowest at middle to high frequencies (1–5 kHz).

A scale that compensates for differences in frequency resolution across the audible range can be useful for gaining insight into the internal respresentation of the spectra of sounds. Such a scale is obtained by expressing a given frequency in terms of the numbers of CBs or ERBs needed to cover the audible range up to that frequency. One scale of this type, based on the traditional CB, is the Bark scale,[8] where the number of barks up to a given frequency is indicated by the symbol z. A good approximation to this scale is[9]

$$z = \frac{26.8}{1 + 1.96/F_c} - 0.53. \tag{3}$$

A scale based on the ERB of the auditory filter and derived from Eq. (2), is[7]

$$\text{Number of ERBs}, E = 21.4\log_{10}(4.37F_c + 1). \tag{4}$$

2.2 The Excitation Pattern

The shape of any given auditory filter characterizes frequency selectivity at a particular center frequency, or,

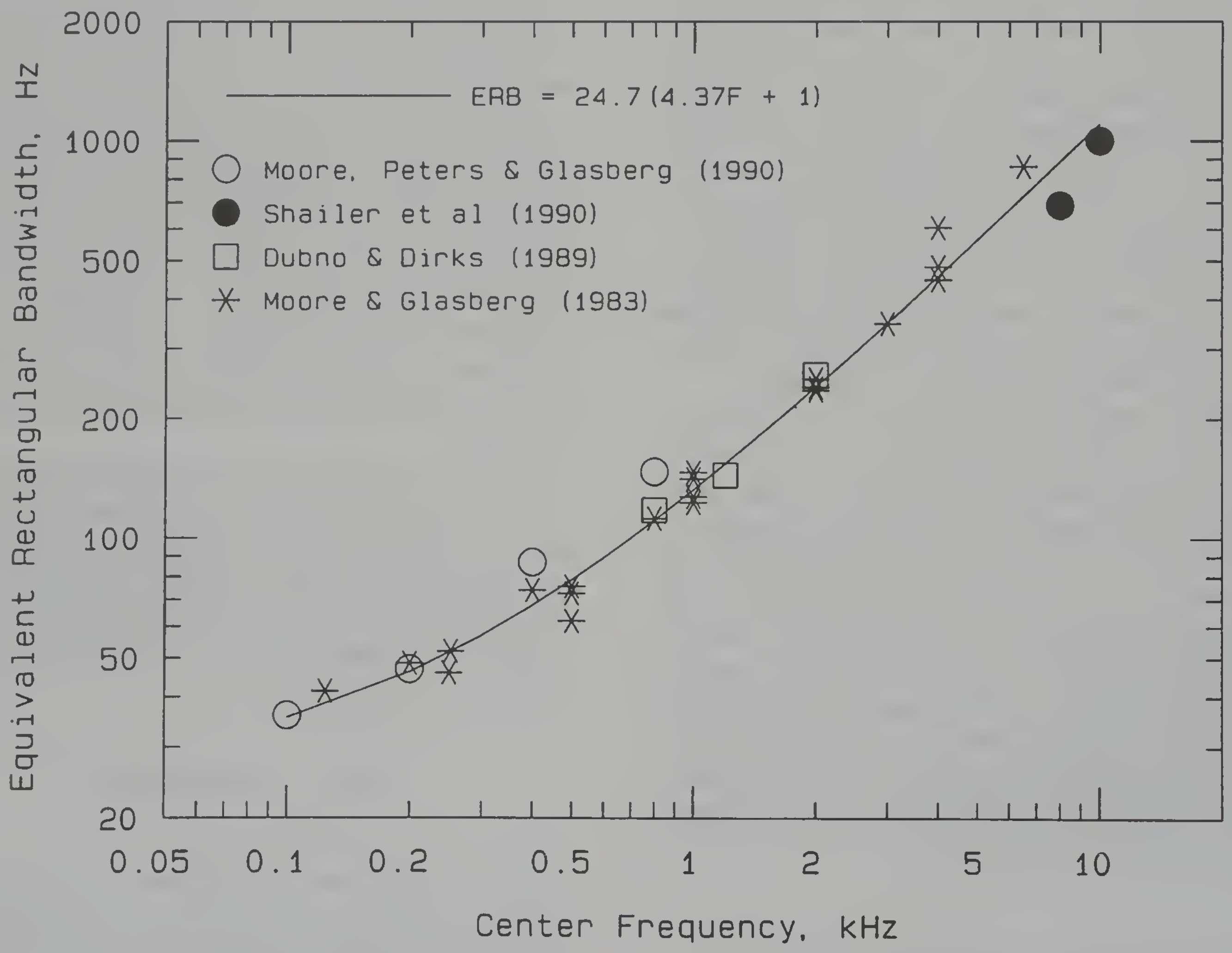


Fig. 1 Estimates of the auditory filter bandwidth from a variety of experiments plotted as a function of center frequency. The solid line represents the equation suggested by Glasberg and Moore.[7]

correspondingly, at a particular place on the basilar membrane. An alternative approach is to characterize the distribution of excitation across frequencies (or places along the basilar membrane) for a given sound: This distribution has been referred to as the excitation pattern of the sound.[10] In terms of the filter-bank analogy, the excitation pattern can be considered as the output of the auditory filters as a function of filter center frequency. Using this definition, it is possible to calculate the excitation pattern of any sound assuming that the auditory filters have the form given in Eq. (1) and have a bandwidth that varies with center frequency according to Eq. (2).[6,7,11] Figure 2 gives an example for a vowel sound. The frequency-related axis is scaled in this figure according to the number of ERBs using Eq. (4). Vowel sounds contain spectral prominences at particular frequencies. These are called formants, and they correspond to resonances in the vocal tract. However, the prominent peaks in the excitation pattern at low center frequencies correspond to individual harmonics, rather than to the first (lowest) formant. Note also that the second, third, and fourth formants, which are clearly seen in the spectrum of the sound, are not clearly separated in the excitation pattern.

2.3 Audibility of Partials in Complex Sounds

When listening to two sinusoids presented simultaneously, either a single sound may be heard, corresponding to the mixture of the two, or two tones may be heard, each with a pitch corresponding to one of the sinuosoids. The minimum frequency separation at which the two sinusoids can be heard separately varies depending on their mean frequency. At 500 Hz it is about 35 Hz, while at 5000 Hz it is about 700 Hz.[12]

When listening to a complex tone containing many components, produced, for example, by a musical instrument, a single pitch is usually heard (see below). However, it is possible to "hear out" individual partials provided attention is directed in an appropriate way, for example, by presenting a comparison sinusoidal tone whose frequency coincides with that of one of the partials.

For multitone complexes, it is generally slightly more difficult to hear the individual components than when only two are present. Particularly for frequencies below 1000 Hz, the frequency separation between adjacent components required to hear out the components is greater than that required for a two-tone complex.[12–14] Figure 3 summarizes results from Plomp[12] and Plomp and Mimpen.[15] It shows the frequency separation between adjacent partials required for a given partial to be heard out from a complex tone with many equal-amplitude components when the components are harmonically related (circles) or nonharmonically related (asterisks). For comparison, the short-dashed line shows the traditional critical-band function and the solid line shows the ERB function described by Eq. (2). Moore and Ohgushi[14] showed that the components in a multicomponent complex had to be separated by about 1.25 ERBs in order for them to be heard out with 75% accuracy. The long-dashed line shows 1.25 times the ERB function; this fits the data rather well.

The harmonics of a periodic complex tone are equally spaced on a linear frequency scale. Thus, for a harmonic complex, the lower harmonics are more easily heard out, or resolved, than the higher harmonics (compare Fig. 2). For a complex tone with equal-amplitude components, only about the first five to eight harmonics are resolvable. This plays an important role in modern theories of pitch perception for complex tones (see below).

2.4 Critical Bandwidth for Loudness

If the total intensity of a complex sound is fixed, its loudness depends on the frequency range over which the sound extends. Consider, as an example, a noise whose total intensity is held constant while the bandwidth is varied. When the bandwidth of the noise is less than a certain value, the loudness is roughly independent of bandwidth. However, as the bandwidth is increased beyond a certain point, the loudness starts to increase. The bandwidth at which loudness starts to increase is known as the critical bandwidth for loudness summation.[16] Its value is slightly greater than the ERB of the auditory filter. The effect is described in more detail in Chapter 118, where its underlying basis is also described.

3 INTRODUCTION TO THE PERCEPTION OF PITCH

Pitch may be defined as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale."[17] In other words, variations in pitch give rise to a sense of melody. Pitch is related to the repetition rate of the waveform of a sound; for a pure tone this corresponds to the frequency, and for a periodic complex tone to the fundamental frequency. There are, however, exceptions to this simple rule. Since pitch is a subjective attribute; it cannot be measured directly. Assigning a pitch value to a sound is generally understood to mean specifying the frequency of a sinusoid having the same subjective pitch as the sound. Sometimes a periodic complex sound, such as a pulse train, is used as a matching stimulus.

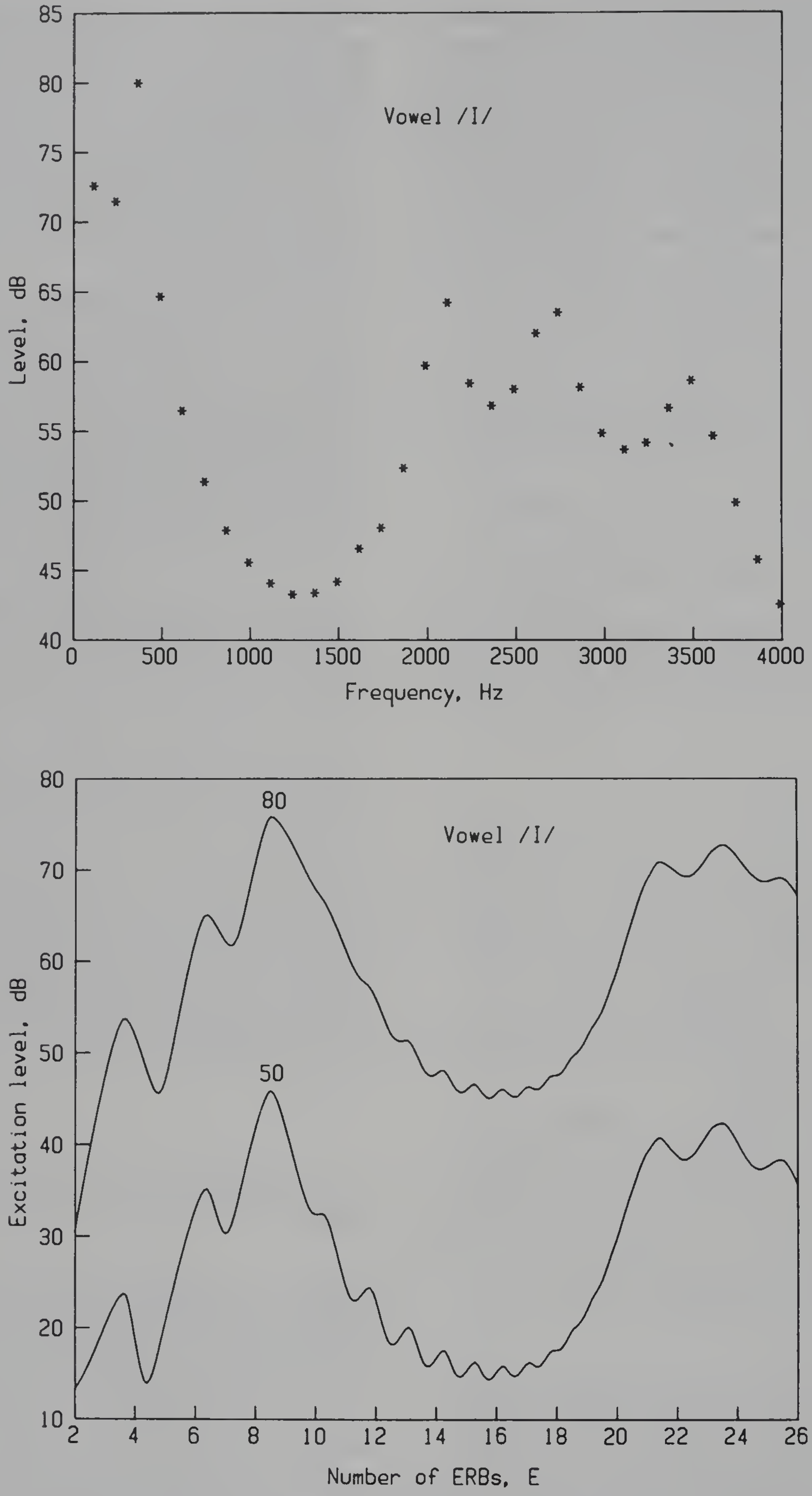


Fig. 2 Spectrum of a vowel sound (top) and the excitation pattern evoked by that sound (bottom) for two different overall sound levels. Note the linear frequency scale in the upper panel and the ERB scale in the lower panel.

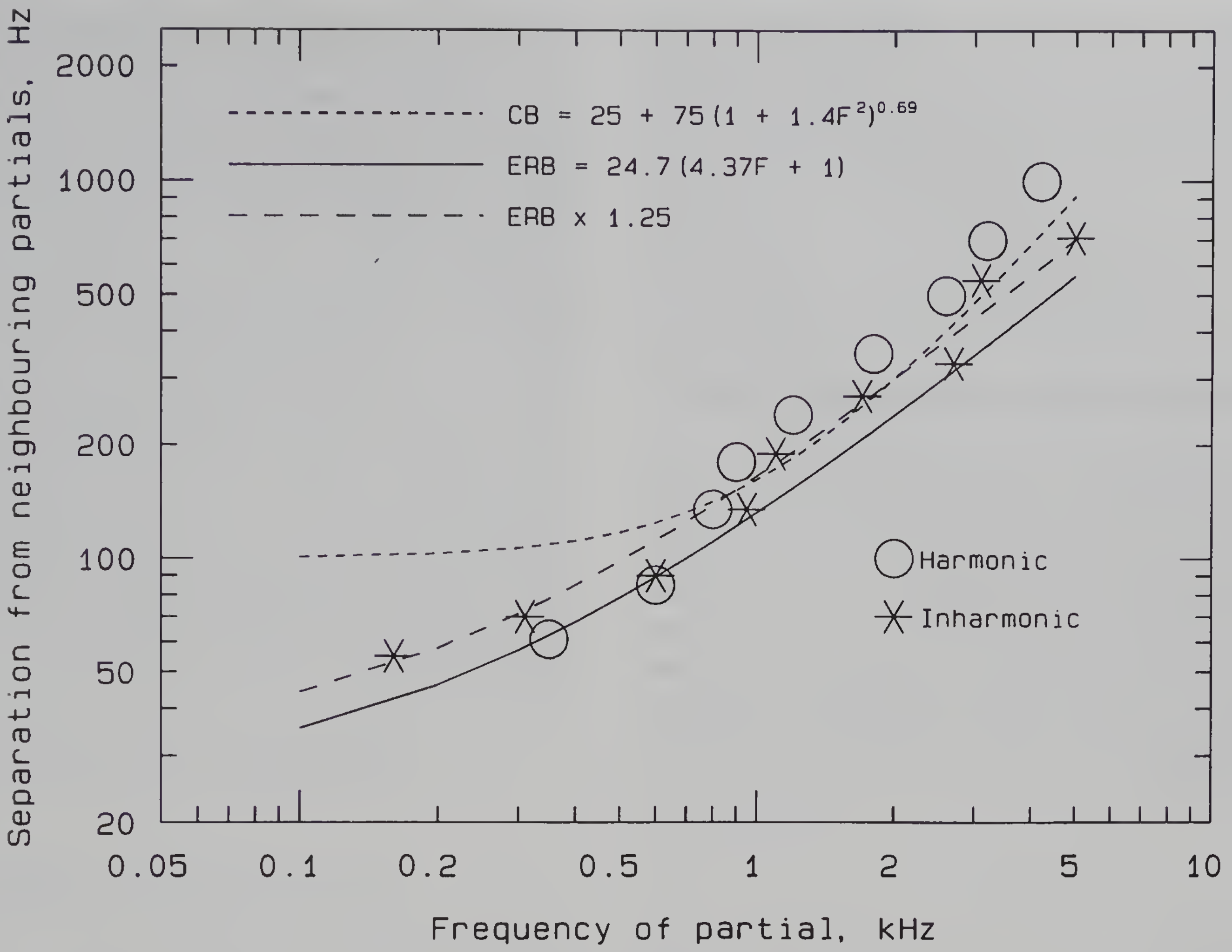


Fig. 3 Frequency separation between adjacent partials required for a given partial to be "heard out" from a complex tone with multiple harmonic (circles) or inharmonic (asterisks) equal-amplitude components, plotted as a function of the frequency of the partial. The short-dashed line shows the traditional CB function. The solid line shows the ERB of the auditory filter as a function of frequency [Eq. (2)]. The long-dashed line shows 1.25 times the ERB.

3.1 Traditional Theories of Pitch Perception

For many years there have been two competing theories of pitch perception. One, the "place" theory of hearing, has two distinct postulates. The first is that the stimulus undergoes a limited-resolution spectral analysis in the inner ear, so that different frequencies (or frequency components in a complex stimulus) excite different places along the basilar membrane. The second is that the pitch of a stimulus is related to the excitation pattern produced by that stimulus; for a pure tone the pitch is generally assumed to correspond to the position of maximum excitation. The first of these two postulates is now well established and has been confirmed in a number of independent ways, including direct observation of the movement of the basilar membrane. The second is still a matter of dispute.

An alternative to the place theory, called the "temporal" theory, suggests that the pitch of a stimulus is related to the time pattern of the neural impulses evoked by that stimulus. Nerve firings tend to occur at a particular phase of the stimulating waveform, and thus the intervals between successive neural impulses approximate integer multiples of the period of the stimulating waveform (this is called phase locking; see Chapter 109). The temporal theory cannot work at very high fundamental frequencies, since phase locking does not occur for frequencies above about 5 kHz. However, the tones produced by musical instruments, the human voice, and most everyday sound sources have fundamental frequencies below this range.

4 PERCEPTION OF THE PITCH OF PURE TONES

This section will discuss the perception and discrimination of the frequency of pure tones. It is important to distinguish between frequency selectivity and frequency discrimination. The former refers to the ability to resolve the frequency components of a complex sound, as discussed earlier. The latter refers to the ability to detect changes in frequency over time. The smallest detectable change is called the difference limen (DL). For place theories, frequency selectivity and frequency discrimination are closely connected; frequency discrimination depends upon the filtering that takes place in the cochlea. For temporal theories, frequency selectivity and frequency discrimination are not necessarily closely connected.

4.1 Frequency Discrimination of Pure Tones

There have been two common ways of measuring frequency discrimination. One measure involves the discrimination of two successive steady tones with slightly different frequencies. This measure will be called the DLF (difference limen for frequency). A second measure uses tones that are frequency modulated at a low rate (typically 2–4 Hz). The amount of modulation required for detection of the modulation is determined. This measure will be called the FMDL (frequency modulation detection limen).

Early studies of frequency discrimination mainly measured FMDLs.[18,19] Recent studies have concentrated more on the measurement of DLFs. A summary of the results of some of these studies is given in Fig. 4, taken from Wier et al.[20] Expressed in hertz the DLF is smallest at low frequencies and increases monotonically with increasing frequency. Expressed as a proportion of center frequency, the DLF tends to be smallest for middle frequencies and larger for very high and very low frequencies. Wier et al. found that the data describing the DLF as a function of frequency fell on a straight line when plotted as log(DLF) versus the square root of frequency; the axes are scaled in this way in Fig. 4. Nelson et al.[21] also found DLFs to be linearly related to the square root of frequency, and they found further that DLFs decrease as the sensation level (SL) of the signal is increased. They at different points along the frequency scale. Their results suggested that DLFs could be described by

$$\log(\mathrm{DLF}) = a\sqrt{f} + k + \frac{m}{\mathrm{SL}}, \tag{5}$$

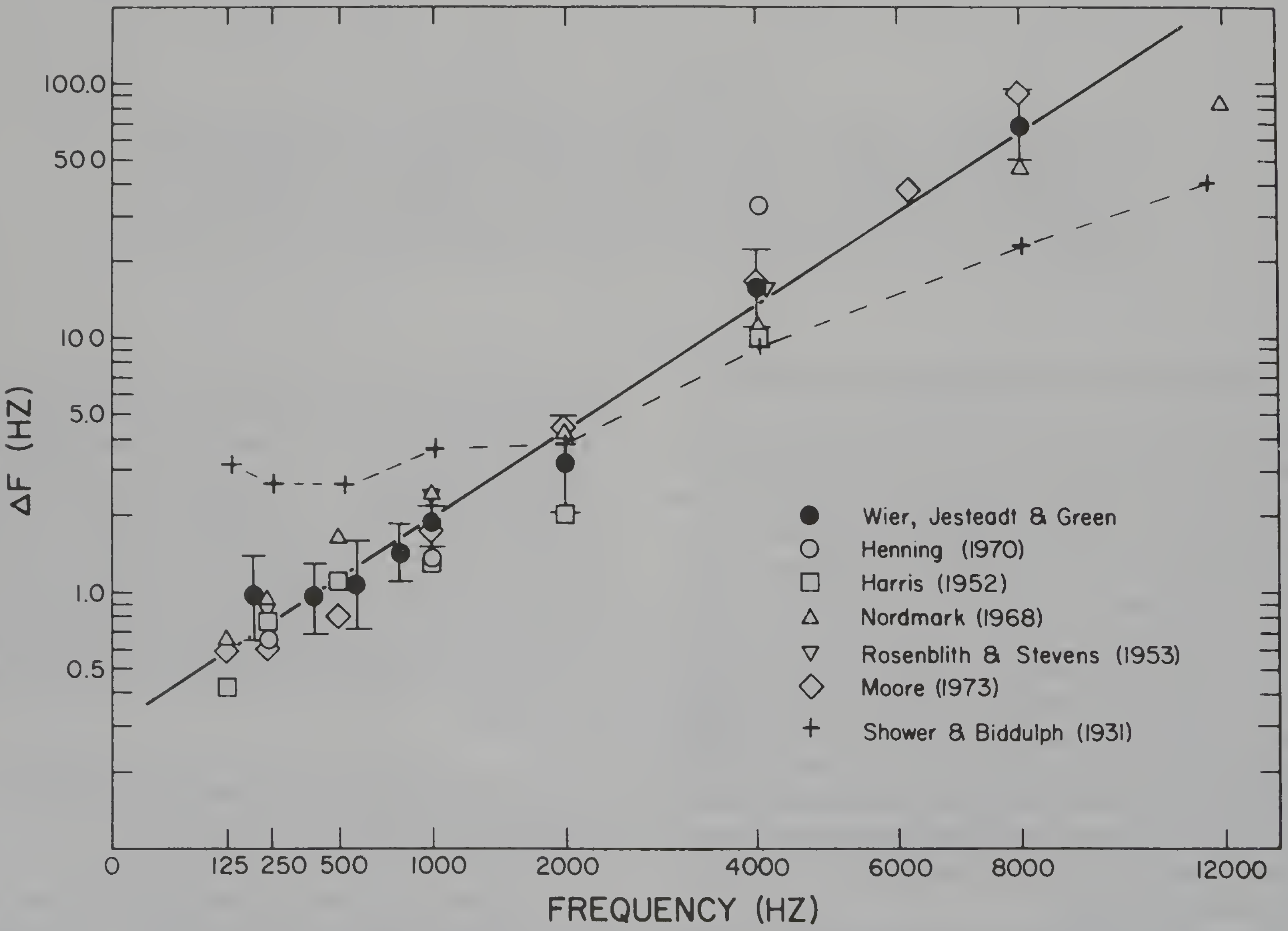


Fig. 4 Summary of the results of several studies measuring frequency DLs. The DLs, ΔF, are plotted in hertz as a function of frequency. All studies measured DLFs, except that of Shower and Biddulph,[18] which measured FMDLs. (From Ref. 20, by permission of the authors and the *Journal of the Acoustical Society of America*.)

where f is the signal frequency in hertz, SL is the sensation level of the signal (in decibels), and a, k, and m are constants. The values of the constants vary somewhat depending on factors such as stimulus duration and the degree of training of the subjects, but typical values are $a = 0.023, k = -0.25$, and $m = 4.3$. The FMDLs, shown as the dashed line in Fig. 4, tend to vary less with frequency than do DLFs.

Place models of frequency discrimination[10,22,23] predict that frequency discrimination should be related to frequency selectivity; the sharper the tuning of peripheral filtering mechanisms, the smaller should be the frequency DL, and the frequency DL should vary with frequency in the same way as the CB. In fact, this prediction fails for DLFs[24,25] but works reasonably well for FMDLs,[10,26] especially when the modulation rate is above about 5 Hz.[27] At low frequencies, DLFs are usually smaller than predicted by place models, but at high frequencies, above 4–5 kHz, they are not. This is consistent with the idea that DLFs are determined mainly by temporal information for frequencies up to about 4–5 kHz, and mainly by place information above that.

4.2 Perception of Musical Intervals

Two tones that are separated in frequency by an interval of one octave (i.e., one has twice the frequency of the other) sound similar. They are judged to have the same name on the musical scale (for example, C or D). This has led several theorists to suggest that there are at least two dimensions to musical pitch. One aspect is related monotonically to frequency (for a pure tone) and is known as *tone height*. The other is related to pitch class (i.e., name note) and is called *tone chroma*.[28]

If subjects are presented with a pure tone of a given frequency f_1 and are asked to adjust the frequency f_2 of a second tone so that it appears to be an octave higher in pitch, they will generally adjust f_2 to be roughly twice f_1. However, when f_1 lies above 2.5 kHz, so that f_2 would lie above 5 kHz, octave matches become very erratic.[29] It appears that the musical interval of an octave is only clearly perceived when both tones are below 5 kHz.

Other aspects of the perception of pitch also change above 5 kHz. A sequence of pure tones above 5 kHz does not produce a clear sense of melody. This has been confirmed by experiments involving musical transposition. For example, Attneave and Olson[30] asked subjects to reproduce sequences of tones (e.g., the NBC chimes) showed an abrupt breakpoint at about 5 kHz above which transposition behavior was erratic. Also, subjects with absolute pitch (the ability to assign name notes without reference to other notes) are very poor at naming notes above 4–5 kHz.[31]

These results are consistent with the idea that the pitch of pure tones is determined by different mechanisms above and below 5 kHz, specifically, by temporal mechanisms at low frequencies and place mechanisms at high frequencies. It appears that the perceptual dimension of tone height persists over the whole audible frequency range, but tone chroma only occurs in the frequency range below 5 kHz.

4.3 Variation of Pitch with Level

The pitch of a pure tone is primarily determined by its frequency. However, sound level also plays a small role. On average, the pitch of tones below about 2000 Hz decreases with increasing level, while the pitch of tones above about 4000 Hz increases with increasing sound level. The early data of Stevens[32] showed rather large effects of sound level on pitch, but more recent data generally show much smaller effects.[33] For tones between 1 and 2 kHz, changes in pitch with level are generally less than 1%. For tones of lower and higher frequencies, the changes can be larger (up to 5%). There are also considerable individual differences both in the size of the pitch shifts with level and in the direction of the shifts. Shifts in pitch with level tend to be very small for complex (musical) tones.

4.4 Diplacusis

When a sinusoid of a given frequency is presented alternately to the two ears, using earphones, it may be perceived as having slightly different pitches in the two ears.[34] The effect can be quantified by asking the listener to adjust the frequency of the tone in one ear so that its pitch matches that of the tone in the other ear. Generally, the shifts measured in this way are less than 1%. This effect has been considered a problem for the temporal theory of pitch. However, adherents of temporal theory argue that the time pattern of neural spikes is analyzed and converted into another form of neural code at a relatively early stage in the auditory system, probably before the point of binaural interaction. If so, then small pitch differences between the two ears could arise at this stage of recoding.

5 PITCH PERCEPTION OF COMPLEX TONES

5.1 Phenomenon of the Missing Fundamental

The classical place theory has difficulty in accounting for the perception of complex tones. For such tones the pitch does not, in general, correspond to the position of maximum excitation on the basilar membrane. A striking illustration of this is provided by the "phenomenon of the missing fundamental." Consider, as an example, a sound consisting of short impulses (clicks) occurring 200 times per second. This sound has a low pitch, which is very close to the pitch of a 200-Hz pure tone, and a sharp timbre. It contains harmonics with frequencies 200, 400, 600, 800, ... Hz. However, it is possible to filter the sound so as to remove the 200-Hz component, and it is found that the pitch does not alter; the only result is a slight change in the timbre of the note. Indeed, all except a small group of midfrequency harmonics can be eliminated, and the low pitch still remains, although the timbre becomes markedly different.

Schouten[35] called the low pitch associated with a group of high harmonics the "residue." Several other names have been used to describe the pitch of a complex tone (whether or not the fundamental component is present), including *periodicity pitch* and *virtual pitch*. This chapter will use the term "virtual pitch." Schouten pointed out that the virtual pitch is distinguishable, subjectively, from a fundamental component that is physically presented or from a fundamental that may be generated (at high sound pressure levels) by nonlinear distortion in the ear. Thus it seems that the perception of a virtual pitch does not require activity at the point on the basilar membrane that would respond maximally to a pure tone of similar pitch. This is confirmed by a demonstration of Licklider[36] that the virtual pitch persists even when low-frequency noise is present that would mask any component at the fundamental frequency. Even when the fundamental component of a complex tone is present, the pitch of the tone is usually determined by harmonics other than the fundamental. Thus the perception of a virtual pitch should not be regarded as unusual. Rather, virtual pitches are normally heard when listening to complex tones.

5.2 Discrimination of the Pitch of Complex Tones

When the waveform repetition rate of a complex tone changes, all of the components change in frequency by the same ratio, and a change in virtual pitch is heard. The ability to detect such changes in pitch is better than the ability to detect changes in a sinusoid at the fundamental frequency,[37] and it can be better than the ability to detect changes in the frequency of any of the sinusoidal components in the complex tone.[38] This indicates that information from the different harmonics is combined or integrated in the determination of virtual pitch. This can lead to very fine discrimination; changes in repetition rate of about 0.2% can often be detected for fundamental frequencies in the range 100–400 Hz.

5.3 Analysis of a Complex Tone in the Peripheral Auditory System

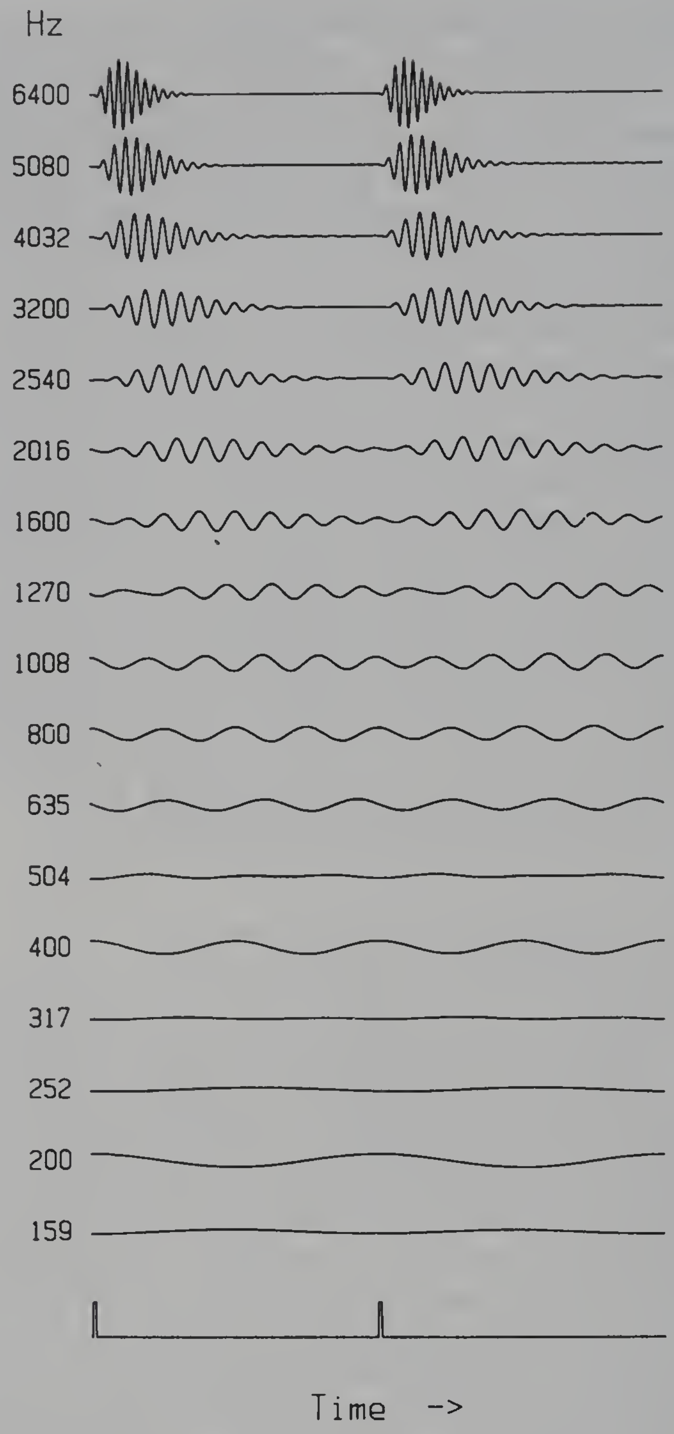


Fig. 5 Simulation of the responses on the basilar membrane to periodic impulses of rate 200 pulses per second. Each number on the left represents the frequency that would maximally excite a given point on the basilar membrane. The waveform that would be observed at that point is plotted opposite that number. (From Ref. 4.)

A simulation of the analysis of a complex tone in the peripheral auditory system is illustrated in Fig. 5. In this example, the complex tone is a periodic pulse train whose spectrum contains many equal-amplitude harmonics. The lower harmonics are partly resolved on the basilar membrane and give rise to distinct peaks in the pattern of activity along the basilar membrane. At a place tuned to the frequency of a low harmonic, the waveform on the basilar membrane is approximately a sine wave at the harmonic frequency. In contrast, the higher harmonics are not resolved and do not give rise to distinct peaks on the basilar membrane. The waveform at places on the membrane responding to higher harmonics is complex and has a repetition rate equal to the fundamental frequency of the sound.

5.4 Theories of Pitch Perception for Complex Tones

Several theories have been proposed to account for virtual pitch. Theories prior to 1980 may be divided into two broad classes. The first, spectral theories, proposed that the perception of the pitch of a complex tone involves two stages. The first stage is an analysis that determines the frequencies of some of the individual sinusoidal components of the complex tone; this depends partly on the resolution of lower harmonics on the basilar membrane. The second stage is a pattern recognizer that determines the pitch of the complex from the frequencies of the resolved components.[39,40] In essence, the pattern recognizer tries to find the harmonic series giving the best match to the resolved frequency components; the fundamental frequency of this harmonic series determines the perceived pitch. For these theories, the lower resolvable harmonics should determine the pitch that is heard.

The alternative, temporal, theories assumed that pitch is based on the time pattern of the waveform at a point on the basilar membrane responding to the higher harmonics. Pitch is assumed to be related to the time interval between corresponding points in the fine structure of the waveform close to adjacent envelope maxima.[41] Nerve firings tend to occur at these points (i.e., phase locking occurs), so this time interval will be present in the time pattern of neural impulses. For such theories, the upper unresolved harmonics should contribute substantially to the pitch that is heard.

Some recent theories (spectrotemporal theories) assume that both frequency analysis (spectral resolution) and time pattern analysis are involved in pitch perception.[4,42–44]

5.5 Physical Variables Influencing Virtual Pitch

Pitch of Inharmonic Complex Tones Schouten et al.[41] investigated the pitch of AM sine waves. If a carrier of frequency f_c is amplitude modulated by a modulator with frequency g, then the modulated wave contains components with frequencies $f_c - g, f_c$, and $f_c + g$. For example, a 2000-Hz carrier modulated 200 times per second contains components at 1800, 2000, and 2200 Hz and has a pitch similar to that of a 200-Hz sine wave.

Consider the effect of shifting the carrier frequency to, say, 2040 Hz. The complex now contains components at 1840, 2040, and 2240 Hz, which do not form a simple harmonic series. The perceived pitch, in this case, corresponds roughly to that of a 204-Hz sinusoid. The shift in pitch demonstrates that the pitch of a complex tone is not determined by the envelope repetition rate, which is 200 Hz, or by the spacing between adjacent components, which is also 200 Hz. In addition, there is an ambiguity of pitch; pitches around 185 Hz and 227 Hz may also be heard.

The perceived pitch can be explained both by spectral theories and by temporal theories. According to the spectral theories, a harmonic complex tone with a fundamental of 204 Hz would provide a good "match"; this would have components at 1836, 2040, and 2244 Hz. Errors in estimating the fundamental occur mainly through errors in estimating the appropriate harmonic number. In the above example the presented components were assumed by the pattern recognizer to be the 9th, 10th, and 11th harmonics of a 204-Hz fundamental. However, a reasonable fit could also be found by assuming the components to be the 8th, 9th, and 10th harmonics of a 226.7-Hz fundamental or the 10th, 11th, and 12th harmonics of a 185.5-Hz fundamental. Thus the theories predict the ambiguities of pitch that are actually observed for this stimulus.

The pitch shift is explained by temporal theories in the following way. For the complex tone with a shifted carrier frequency, the time intervals between corresponding peaks in the fine structure of the wave are slightly less than 5 ms; roughly 10 periods of the carrier occur for each period of the envelope, but the exact time for 10 periods of a 2040-Hz carrier is 4.9 ms, which corresponds to a pitch of 204 Hz. This is illustrated in Fig. 6. Nerve spikes can occur at any of the prominent peaks in the waveform, labeled 1, 2, 3 and $1', 2', 3'$. Thus the time intervals between successive nerve spikes fall into several groups, corresponding to the intervals 1–1′, 1–2′, 1–3′, 2–1′, ... This can account for the fact that the pitches of these stimuli are ambiguous. The intervals 1–1′, 2–2′, and 3–3′ correspond to the most prominent pitch of about 204 Hz, while other intervals, such as 1–2′ and 2–1′, correspond to the other pitches that may be perceived.

Existence Region of Virtual Pitch According to the spectral theories, a virtual pitch should only be heard when at least one of the sinusoidal components of the stimulus can be heard out. Thus a virtual pitch should not be heard for stimuli containing only very high unresolvable harmonics. Ritsma[45] investigated the audibility of virtual ptiches for AM sinusoids containing three components as a function of the modulation rate and the harmonic number (carrier frequency divided by modulation rate). Later,[46] he extended the results to complexes with a greater number of components. He found that the tonal character of the virtual pitch existed only within a limited frequency region, referred to as the "existence region."

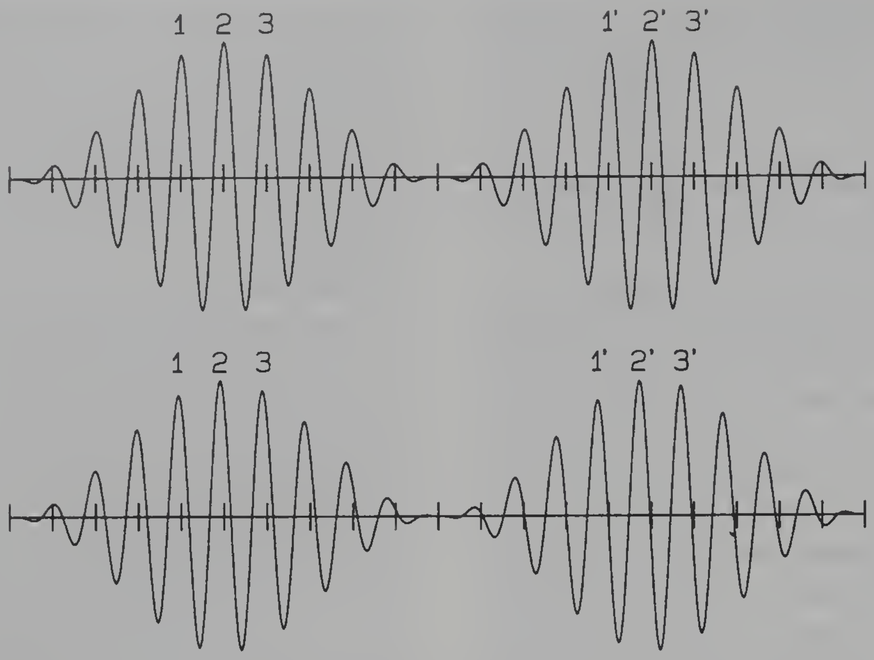


Fig. 6 Waveforms of two amplitude-modulated sinusoids. In the upper waveform, the carrier frequency (2000 Hz) is an exact multiple of the modulation frequency (200 Hz). In the lower waveform, the carrier frequency is shifted slightly upward, to 2040 Hz.

When the harmonic number was too high (above about 20), only a kind of high buzz of undefinable pitch was heard. However, virtual pitches could clearly be heard for harmonic numbers between 10 and 20, provided the frequencies of the harmonics were below about 5 kHz. These harmonics would not be resolvable.

Ritsma's results may have been influenced by the presence of combination tones in the frequency region below the lowest components in the complex tones, as has since been pointed out by Ritsma[47] himself. (See Chapter 108 for more information on combination tones.) However, more recent experiments, which have made use of background noise to mask combination tones, also indicate that virtual pitch can be perceived when no resolvable harmonics are present. Moore and Rosen[48] and Houtsma and Smurzynski[49] have demonstrated that stimuli containing only high, unresolvable harmonics can evoke a sense of musical interval and hence of pitch. The spectral theories cannot account for this.

Principle of Dominance Ritsma[50] carried out an experiment to determine which components in a complex sound are most important in determining its pitch. He presented complex tones in which the frequencies of a small group of harmonics were multiples of a fundamental that was slightly higher or lower than the fundamental of the remainder. The subject's pitch judgments were used to determine whether the pitch of the complex as a whole was affected by the shift in the group of harmonics. Ritsma found that "for fundamental frequencies in the range 100 Hz to 400 Hz, and for sensation levels up to at least 50 dB above threshold of the entire signal, the frequency band consisting of the third, fourth and fifth harmonics tends to dominate the pitch senstation as long as its amplitude exceeds a minimum absolute level of about 10 dB above threshold."

This finding has been broadly confirmed in other ways,[51] although the data of Moore et al.[38,52] show that there are large individual differences in which harmonics are dominant, and for some subjects the first two harmonics play an important role. Other data also show that the dominant region is not fixed in terms of harmonic number but also depends somewhat on absolute frequency.[51,53] For high fundamental frequencies (above about 1000 Hz), the fundamental is usually the dominant component, while for very low fundamental frequencies (around 50 Hz), harmonics above the fifth may be dominant.[54]

On the whole, these results are better explained by the spectral theories. The third, fourth, and fifth harmonics are relatively well resolved on the basilar membrane and are usually separately perceptible. However, for very low fundamental frequencies (around 50 Hz), harmonics in the dominant region (above the fifth harmonic) would not be resolvable.

Pitch of Dichotic Two-Tone Complexes Houtsma and Goldstein[55] investigated the perception of stimuli that comprised just two harmonics with successive harmonic numbers, for example, the fourth and fifth harmonics of a (missing) fundamental. They found that musically trained subjects were able to identify musical intervals between such stimuli when they were presented dichotically (one harmonic to each ear), presumably indicating that virtual pitches were heard. These results cannot be explained by temporal theories, which require an interaction of components on the basilar membrane. They imply that "the pitch of these complex tones is mediated by a central processor operating on neural signals derived from those effective stimulus harmonics which are tonotopically resolved."[55] However, it is not yet entirely clear that the findings apply to the pitches of complex tones in general. The pitch of a two-tone complex is not very clear; indeed, many observers do not hear a single low pitch but rather perceive the component tones individually.[56]

Pitch of Stimuli without Spectral Cues A number of workers have investigated pitches evoked by stimuli that contain no spectral peaks and therefore would not produce a well-defined maximum on the basilar membrane. Such pitches cannot arise on the basis of place information and so provide evidence for the operation of temporal mechanisms.

Miller and Taylor[57] used random white noise that was turned on and off abruptly and periodically at various rates. Such interrupted noise has a flat long-term magnitude spectrum. They reported that the noise had a pitchlike quality for interruption rates between 100 and 250 Hz. Burns and Viemeister[58,59] showed that noise that was sinsusoidally amplitude modulated had a pitch corresponding to the modulation rate. As is the case for interrupted noise, the long-term magnitude spectrum of white noise remains flat after modulation. They found that the pitch could be heard for modulation rates up to 800–1000 Hz and that subjects could recognize musical intervals and melodies played by varying the modulation rate.

It should be noted that, although the long-term magnitude spectrum of interrupted or amplitude-modulated noise is flat, some spectral cues exist in the short-term magnitude spectrum.[69] Thus, the use of place cues to extact the pitch of these stimuli cannot be completely ruled out. This is not the case, however, for experiments showing pitch sensations as a result of the binaural interaction of noise stimuli. Huggins and Cramer[61] fed white noise from the same noise generator to both ears via headphones. The noise to one ear went through an all-pass filter, which passed all audible frequency components without change of amplitude but produced a phase change in a small frequency region. A faint pitch was heard corresponding to the frequency region of the phase transition, although the stimulus presented to each ear alone was white noise and produced no pitch sensation. The effect was only heard when the phase change was in the frequency range below about 1 kHz. A phase shift of a narrow band of the noise results in that band being heard in a different position in space from the rest of the noise. This spatial separation presumably produces the pitchlike character associated with that narrow band. Pitches produced by binaural interaction have also been reported by Fourcin[62] and Bilsen and Goldstein.[63]

Effect of Relative Phases of Components on Pitch Changes in the relative phases of the components making up a complex tone change the temporal structure of the tone but have little effect on the auditory representation of the lower, resolved harmonics. Thus, spectral theories predict that relative phase should not influence pitch, while temporal theories predict that relative phase should affect pitch, pitch being more distinct for phase relations producing "peaky" waveforms after auditory filtering (waveforms with high crest factors). For tones containing many harmonics, phase has only small effects on the value of the pitch heard,[64] but it does affect the clarity of pitch.[65] For tones containing only a few high harmonics, phase can affect both the pitch value and the clarity of pitch.[66] Phase can affect the discrimination of changes in virtual pitch, and the phase effects are larger in hearing-impaired subjects than in normal subjects, presumably because auditory filters are broader in impaired subjects and produce more interaction of the components in the peripheral auditory system.[54] Overall, these results suggest that the temporal structure of complex sounds does influence pitch perception.

6 SPECTROTEMPORAL MODELS OF PITCH PERCEPTION OF COMPLEX TONES

The evidence reviewed above indicates that neither spectral nor temporal theories can account for all aspects of the pitch perception of complex tones. Spectral theories cannot account for the following facts: A virtual pitch can be perceived when only high, unresolvable harmonics are present; stimuli without spectral peaks can give rise to a pitch sensation; and the relative phases of the components can affect pitch perception. Temporal theories cannot account for these facts: The lower, resolvable harmonics are dominant in determining pitch; and a virtual pitch may be heard when there is no possibility of the components interacting in the peripheral auditory system.

These findings have led several workers to propose theories in which both spectral and temporal mechanisms play a role; the initial place/spectral analysis in the cochlea is followed by an analysis of the time pattern of the waveform at each place or of the neural spikes evoked at each place.[4,42–44] The model proposed by Moore is based on the analysis of interspike intervals at each characteristic frequency (CF) and includes a device that compares the time intervals present at different CFs and searches for common time intervals. The device may also integrate information over time. In general the time interval that is found most often corresponds to the period of the fundamental component, and the perceived pitch corresponds to the reciprocal of this interval. Spectrotemporal theories can account for most existing data on the pitch perception of complex tones and so are more satisfactory than either spectral or temporal theories.

Acknowledgments

I thank David Green, Adrian Houtsma, Dennis McFadden and Dixon Ward for helpful comments on an earlier version of this chapter.

REFERENCES

1. J. O. Pickles, "The Neurophysiological Basis of Frequency Selectivity," in B. C. J. Moore (Ed.), *Frequency Selectivity in Hearing*, Academic, London, 1986, pp. 51–121.
2. H. Fletcher, "Auditory Patterns," *Rev. Mod. Phys.*, Vol. 12, 1940, pp. 47–65.
3. B. Scharf, "Critical Bands," in J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory*, Academic, New York, 1970, pp. 157–202.
4. B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 3rd ed., Academic, London, 1989.
5. R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy, "The Deterioration of Hearing with Age: Frequency Selectivity, the Critical Ratio, and Audiogram, and Speech Threshold," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 1788–1803.
6. B. C. J. Moore and B. R. Glasberg, "Formulae Describing Frequency Selectivity as a Function of Frequency and Level and Their Use in Calculating Excitation Patterns," *Hear. Res.*, Vol. 28, 1987, pp. 209–225.
7. B. R. Glasberg and B. C. J. Moore, "Derivation of Auditory Filter Shapes from Notched-Noise Data," *Hearing. Res.*, Vol. 47, 1990, pp. 103–138.
8. E. Zwicker and E. Terhardt, "Analytical Expressions for Critical Band Rate and Critical Bandwidth as a Function of Frequency," *J. Acoust. Soc. Am.*, Vol. 68, 1980, pp. 1523–1525.
9. H. Traunmuller, "Analytical Expressions for the Tonotopic Sensory Scale," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 97–100.
10. E. Zwicker, "Masking and Psychological Excitation as Consequences of the Ear's Frequency Analysis," in R. Plomp and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*, Sijthoff, Leiden 1970, pp. 376–394.
11. B. C. J. Moore and B. R. Glasberg, "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," *J. Acoust. Soc. Am.*, Vol. 74, 1983, pp. 750–753.
12. R. Plomp, "The Ear as a Frequency Analyzer," *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 1628–1636.
13. B. C. J. Moore, "Some Experiments Relating to the Perception of Complex Tones," *Q. J. Exp. Psych.*, Vol. 25, 1973, pp. 451–475.
14. B. C. J. Moore and K. Ohgushi, "Audibility of Partials in Inharmonic Complex Tones," *J. Acoust. Soc. Am.*, Vol. 93, 1993, pp. 452–461.
15. R. Plomp and A. M. Mimpen, "The Ear as a Frequency Analyzer II," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 764–767.
16. E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical Bandwidth in Loudness Summation," *J. Acoust. Soc. Am.*, Vol. 29, 1957, pp. 548–557.
17. American Standards Association, "Acoustical Terminology," SI, 1-1960, American Standards Association, New York, 1960.
18. E. G. Shower and R. Biddulph, "Differential Pitch Sensitivity of the Ear," *J. Acoust. Soc. Am.*, Vol. 2, 1931, pp. 275–287.
19. E. Zwicker, "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs," *Acustica*, Vol. 6, 1956, pp. 356–381.
20. C. C. Wier, W. Jesteadt, and D. M. Green, "Frequency Discrimination as a Function of Frequency and Sensation Level," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 178–184.
21. D. A. Nelson, M. E. Stanton, and R. L. Freyman, "A General Equation Describing Frequency Discrimination as a Function of Frequency and Sensation Level," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 2117–2123.
22. G. B. Henning, "A Model for Auditory Discrimination and Detection," *J. Acoust. Soc. Am.*, Vol. 42, 1967, pp. 1325–1334.
23. W. M. Siebert, "Frequency Discrimination in the Auditory System: Place or Periodicity Mechanisms," *Proc. IEEE*, Vol. 58, 1970, pp. 723–730.
24. B. C. J. Moore, "Relation between the Critical Bandwidth and the Frequency-Difference Limen," *J. Acoust. Soc. Am.*, Vol. 55, 1974, p. 359.
25. B. C. J. Moore and B. R. Glasberg, "The Role of Frequency Selectivity in the Perception of Loudness, Pitch and Time," in B. C. J. Moore (Ed.), *Frequency Selectivity in Hearing*, Academic, London 1986, pp. 251–308.
26. B. C. J. Moore and B. R. Glasberg, "Mechanisms Underlying the Frequency Discrimination of Pulsed Tones and the Detection of Frequency Modulation," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 1722–1732.
27. A. Sek and B. C. J. Moore, "Frequency Discrimination as a Function of Frequency, Measured in Several Ways," *J. Acoust. Soc. Am.*, Vol. 97, 1995, pp. 2479–2486.
28. A. Bachem, "Tone Height and Tone Chroma as Two Different Pitch Qualities," *Acta Psychol.*, Vol. 7, 1950, pp. 80–88.

29. W. D. Ward, "Subjective Musical Pitch," *J. Acoust. Soc. Am.*, Vol. 26, 1954, pp. 369–380.
30. F. Attneave and R. K. Olson, "Pitch as a Medium: A New Approach to Psychophysical Scaling," *Am. J. Psychol.*, Vol. 84, 1971, pp. 147–166.
31. K. Ohgushi and T. Hatoh, "Perception of the Musical Pitch of High Frequency Tones," in Y. Cazals, L. Demany, and K. Horner (Eds.), *Ninth International Symposium on Hearing: Auditory Physiology and Perception*, Pergamon, Oxford, 1991, pp. 207–212.
32. S. S. Stevens, "The Relation of Pitch to Intensity," *J. Acoust. Soc. Am.*, Vol. 6, 1935, pp. 150–154.
33. J. Verschuure and A. A. Van Meeteren, "The Effect of Intensity on Pitch," *Acustica*, Vol. 32, 1975, pp. 33–44.
34. G. Van den Brink, "Two Experiments on Pitch Perception: Diplacusis of Harmonic AM Signals and Pitch of Inharmonic AM Signals," *J. Acoust. Soc. Am.*, Vol. 48, 1970, pp. 1355–1365.
35. J. F. Schouten, "The Residue Revisited," in R. Plomp and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*, Sijthoff, Lieden, The Netherlands 1970, pp. 41–54.
36. J. C. R. Licklider, "Auditory Frequency Analysis," in C. Cherry (Ed.), *Information Theory*, Academic, New York, 1956, pp. 253–268.
37. J. L. Flanagan and M. G. Saslow, "Pitch Discrimination for Synthetic Vowels," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 435–442.
38. B. C. J. Moore, B. R. Glasberg, and M. J. Shailer, "Frequency and Intensity Difference Limens for Harmonics within Complex Tones," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 550–561.
39. J. L. Goldstein, "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones," *J. Acoust. Soc. Am.*, Vol. 54, 1973, pp. 1496–1516.
40. E. Terhardt, "Pitch, Consonance, and Harmony," *J. Acoust. Soc. Am.*, Vol. 55, 1974, pp. 1061–1069.
41. J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, "Pitch of the Residue," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 1418–1424.
42. P. Srulovicz and J. L. Goldstein, "A Central Spectrum Model: A Synthesis of Auditory-Nerve Timing and Place Cues in Monaural Communication of Frequency Spectrum," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 1266–1276.
43. R. D. Patterson, "A Pulse Ribbon Model of Monaural Phase Perception," *J. Acoust. Soc. Am.*, Vol. 82, 1987, pp. 1560–1586.
44. R. Meddis and M. Hewitt, "Virtual Pitch and Phase Sensitivity Studied Using a Computer Model of the Auditory Periphery: Pitch Identification," *J. Acoust. Soc. Am.*, Vol. 89, 1991, pp. 2866–2882.
45. R. J. Ritsma, "Existence Region of the Tonal Residue. I," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 1224–1229.
46. R. J. Ritsma, "Existence Region of the Tonal Residue. II," *J. Acoust. Soc. Am.*, Vol. 35, 1963, pp. 1241–1245.
47. R. J. Ritsma, "Periodicity Detection," in R. Plomp and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*, Sijthoff, Leiden, 1970, pp. 250–263.
48. B. C. J. Moore and S. M. Rosen, "Tune Recognition with Reduced Pitch and Interval Information," *Q. J. Exper. Psychol.*, Vol. 31, 1979, pp. 229–240.
49. A. J. M. Houtsma and J. Smurzynski, "Pitch Identification and Discrimination for Complex Tones with Many Harmonics," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 304–310.
50. R. J. Ritsma, "Frequencies Dominant in the Perception of the Pitch of Complex Sounds," *J. Acoust. Soc. Am.*, Vol. 42, 1967, pp. 191–198.
51. R. Plomp, "Pitch of Complex Tones," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 1526–1533.
52. B. C. J. Moore, B. R. Glasberg, and R. W. Peters, "Relative Dominance of Individual Partials in Determining the Pitch of Complex Tones," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 1853–1860.
53. R. D. Patterson and F. L. Wightman, "Residue Pitch as a Function of Component Spacing," *J. Acoust. Soc. Am.*, Vol. 59, 1976, pp. 1450–1459.
54. B. C. J. Moore and R. W. Peters, "Pitch Discrimination and Phase Sensitivity in Young and Elderly Subjects and Its Relationship to Frequency Selectivity," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 2881–2893.
55. A. J. M. Houtsma and J. L. Goldstein, "The Central Origin of the Pitch of Pure Tones: Evidence from Musical Interval Recognition," *J. Acoust. Soc. Am.*, Vol. 51, 1972, pp. 520–529.
56. G. F. Smoorenburg, "Pitch Perception of Two-Frequency Stimuli," *J. Acoust. Soc. Am.*, Vol. 48, 1970, pp. 924–941.
57. G. A. Miller and W. Taylor, "The Perception of Repeated Bursts of Noise," *J. Acoust. Soc. Am.*, Vol. 20, 1948, pp. 171–182.
58. E. M. Burns and N. F. Viemeister, "Nonspectral Pitch," *J. Acoust. Soc. Am.*, Vol. 60, 1976, pp. 863–869.
59. E. M. Burns and N. F. Viemeister, "Played Again SAM: Further Observations on the Pitch of Amplitude-Modulated Noise," *J. Acoust. Soc. Am.*, Vol. 70, 1981, pp. 1655–1660.
60. J. R. Pierce, R. Lipes, and C. Cheetham, "Uncertainty Concerning the Direct Use of Time Information in Hearing: Place Clues in White Spectra Stimuli," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 1609–1621.
61. W. H. Huggins and E. M. Cramer, "Creation of Pitch through Binaural Interaction," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 413–417.
62. A. J. Fourcin, "Central Pitch and Auditory Lateralization," in R. Plomp and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*, Sijthoff, Leiden, 1970, pp. 319–328.
63. F. A. Bilsen and J. L. Goldstein, "Pitch of Dichotically Delayed Noise and Its Possible Spectral Basis," *J. Acoust. Soc. Am.*, Vol. 55, 1974, pp. 292–296.

64. R. D. Patterson, "The Effects of Relative Phase and the Number of Components on Residue Pitch," *J. Acoust. Soc. Am.*, Vol. 53, 1973, pp. 1565–1572.
65. C. Lundeen and A. M. Small, "The Influence of Temporal Cues on the Strength of Periodicity Pitches," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 1578–1587.
66. B. C. J. Moore, "Effects of Relative Phase of the Components on the Pitch of Three-Component Complex Tones," in E. F. Evans and J. P. Wilson (Eds.), *Psychophysics and Physiology of Hearing*, Academic, London, 1977, pp. 349–358.

117

FUNCTIONS OF THE BINAURAL SYSTEM

ERVIN R. HAFTER AND CONSTANTINE TRAHIOTIS

The binaural nervous system combines information from the two ears. This allows the listener to localize the directions of sounds and, in some cases, to orient and navigate without the use of vision. When there are reflective surfaces, binaural hearing may be useful for computation of the distance to an acoustic source and, through analysis of reverberation, for enrichment of the auditory experience, as in concert halls. In a background of maskers or distractors, binaural processing can enhance the detectability of signals and improve the intelligibility of speech. The discussion here will concentrate first on the binaural information available in open sound fields and then move to results from experiments that have employed earphones to investigate the binaural cues individually and in combination. Numerous excellent reviews[1–7] present these topics in greater detail.

1 SPATIAL RELATIONS AND DIFFERENCES IN THE ACOUSTIC PATHWAYS

1.1 Common Schemes Describing Three-Dimensional Auditory Space

Three such methods are shown in Fig. 1. Each assumes that the source lies on the surface of an imaginary sphere that is concentric with the listener's head. The line through the ears is called the *aural axis*. A *horizontal plane* containing the aural axis divides auditory space into upper and lower halves. Sources above or below this plane are said to have *elevation* or *declination*. Orthogonal to the aural axis is the *midsagittal* or *median plane* that passes vertically through the center of the head. It divides auditory space into left and right halves, and sources away from this plane are said to have *lateral* or *azimuthal displacement*. For these sources, differences in the acoustic paths produce the *interaural differences* between sounds at the two ears that are the bases for binaural processing.

The geodesic description shown in Fig. 1*a*[4] imagines a right triangle whose hypotenuse connects the center of the head to the source and whose altitude is the vertical projection from the source to the horizontal plane. Lateral displacement is defined by the angle ϕ in the horizontal plane between the base of the triangle and the median plane. The locus of all displacements with the same ϕ is then the geographer's longitude. Elevation is defined as the angle δ between the hypotenuse of the triangle and the horizontal plane. The locus of constant elevation is the geographer's latitude. A weakness of this system is that, for a constant lateral displacement, the differences between pathways to the two ears grow smaller with increased elevation, reaching zero at $\delta = 90°$. To correct this, the scheme in Fig. 1*b*[8] describes lateral displacement by an angle (α) between the aural axis and a radius from the center of the head to the source while elevation is described by the angle (β) between the horizontal plane and a plane through the aural axis and the source. Note that the locus of points for a single value of α is a circle parallel to the median plane, making the differences between pathways to the two ears the same regardless of elevation. An argument against this system is that β does not capture the implied meaning of elevation as well as ϕ in Fig. 1*a* because the value of β for a given height above the horizontal plane depends on α. Another system in common use is the *double-pole coordinate system*[9,10] shown in Fig. 1*c*. Here, azimuth and elevation are defined by angles between a radius through the source and the median plane and horizontal planes,

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

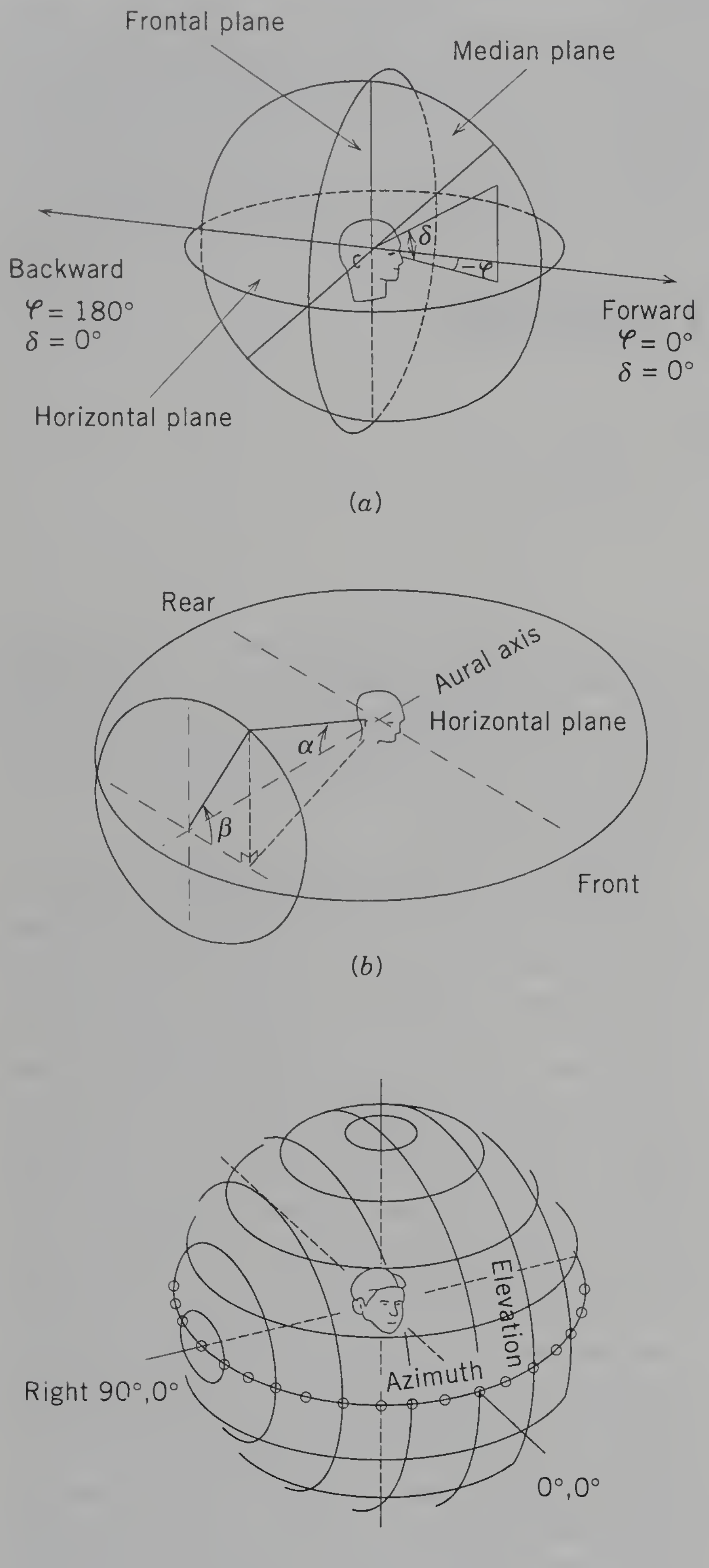


Fig. 1 Three different descriptions of auditory space: (*a*) geodesic[4]; (*b*) with constant binaural distances for a single lateral displacement[8]; (*c*) bipolar.[10]

respectively. In this case, the locus of points defined by a single *azimuth* is a circle parallel to the median plane while the locus defined by a single *elevation* is a circle parallel to the horizontal plane.

1.2 Binaural Cone of Confusion

The simplifying assumption that the head is well modeled as a sphere[11] makes the difference between path lengths identical for all locations along one of the vertical circles described by a value of α in Fig. 1*b* or azimuth in Fig. 1*c*. Extending this *circle of ambiguity* in distance produces what is generally called a binaural *cone of confusion*, descriptive of all points in three-dimensional space whose binaural pathways are identical. Were there no other information, sources along this cone would be indistinguishable from one another.

2 EFFECTS OF LATERAL DISPLACEMENT

2.1 Interaural Time-Differences (ITD)

One of the essential cues for lateral displacement comes from the extra time needed for a stimulus to reach the further ear. This difference in time between signals at the two ears is called the ITD. The ear facing the source, often called the *proximal ear*, is stimulated by incident and scattered sound, including sounds reflected from the torso. The more distant, or *distal*, ear often falls into the acoustic shadow cast by the head, where it receives only the sum of "creeping waves" (assuming negligible scatter from the torso) created by diffraction at the "shadow boundary." Figure 2 shows calculated as well as measured values of the ITD obtained from microphones placed in the ear canals of a KEMAR (Knowles Electronics Manikin for Acoustic Research) manikin based on tones presented from several different azimuths in the horizontal plane.[12] Each creeping wave has its own attenuation and phase velocity. While this velocity approaches the ambient speed of sound for frequencies above about 2000 Hz, it decreases considerably with frequency. Thus Fig. 2 shows that IDTs are greatest with low frequencies but remain essentially constant above 2000 Hz where the IDT is well described by the assumption of a spherical head (Ref. 1; for discussion see Ref. 13). The maximum value of ITD produced by a source on the aural axis is determined by the width of the human head; above 2000 Hz this is about 410 μs at birth,[14] rising to about 680 μs at maturity.[15] Small perturbations in the ITD at higher frequencies result from the complex geometry of the head and pinnas as well as reflections from the torso. Figure 3 contains a summary of ITDs compiled from several sources[16,17] and plotted as a function of azimuth.

2.2 Interaural Sound Pressure Level Differences

The other primary cue for lateral displacement is the difference in sound pressure level between stimuli at the

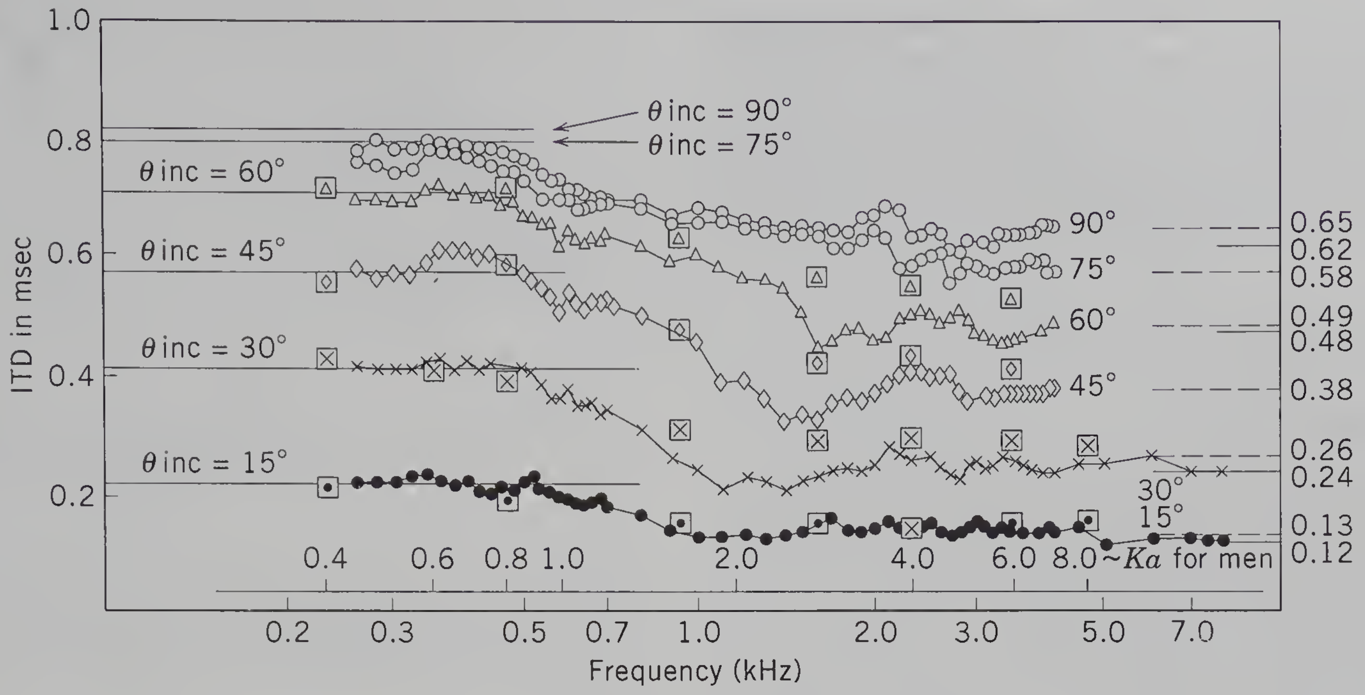


Fig. 2 The ITDs obtained from the ear canals of a KEMAR for sounds at different azimuths in the horizontal plane.[12]

two ears. The interaural sound pressure levels (ILDs) arise primarily from the acoustic shadow cast by the listener's own head onto the distal ear. The ILDs depend heavily on the stimulus frequency. Figure 4 presents ILDs for three frequencies as a function of azimuth and for many frequencies at an azimuth of 90°. The data were derived[2] from a wide-ranging synthesis of results from many studies.[16,17] The upper plot in (*a*) shows that the ILD is generally larger with greater azimuths and higher frequencies. However, because the sounds at each ear are summations of waves arriving along multiple paths, the complex geometry of the head and pinnas as well as reflections from the torso can lead to irregularities; an

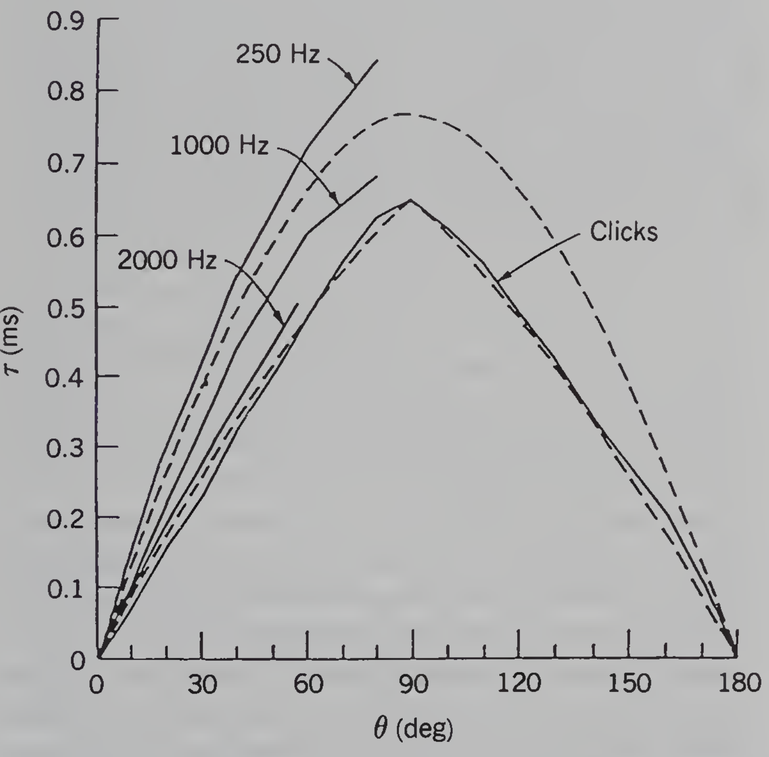


Fig. 3 The ITDs measured as a function of azimuth (redrawn[2] compilation of multiple sources[16,17]).

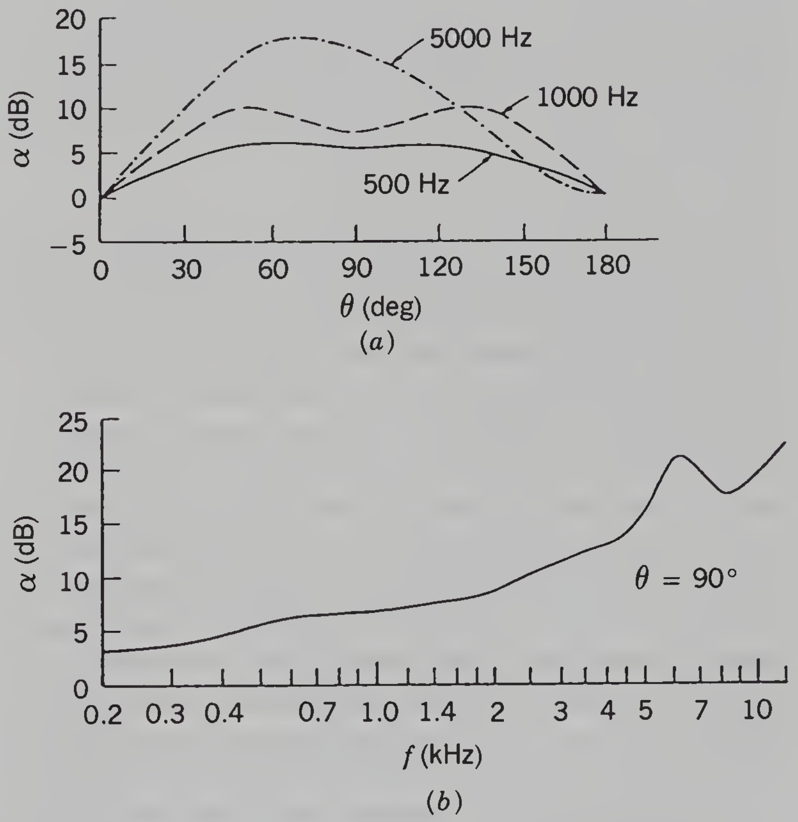


Fig. 4 The ILDs plotted for three frequencies as a function of azimuth and an azimuth of 90° as a function of frequency (redrawn[2] compilation of multiple sources[16,17]).

example of this is the case of the azimuthal region from 130° and 180° where the ILD is actually less for 5000 Hz than for 1000 Hz. Similarly, the lower plot in (*b*) shows that the increase in ILD with frequency is roughly monotonic with the exception again of complexities at high frequencies. (For greater detail, see Refs. 10, 13, 16, and 17.)

3 FREE-FIELD LOCALIZATION

3.1 Duplex Theory

It has long been recognized[18] that there are potential problems with the use of each of the basic interaural cues for localization. First consider the extraction of ITD from an ongoing, narrow-band signal. As shown in Fig. 2, the ITD can be as large as 0.65 ms for a source lying along the aural axis. This offers no problem for low frequencies, but at high frequencies, the ITD can exceed the signal's period and, in so doing, lead to ambiguities in localization. To illustrate the problem, Fig. 5 shows the effects of displacing the source laterally to the left of center. The leading and lagging waveforms at the left and right ears are depicted by solid and dashed curves, respectively, while the ITD is represented by the difference between the dotted, vertical lines. For a periodic signal such as a sinusoid, the ITD may be derived from any point along the ongoing signal by calculating ITD $= \theta/2\pi f$, with θ being the interaural phase delay (in radians) and f being the frequency. For a low frequency, such as that in Fig. 5*a*, the signal's period is more than twice the maximum possible ITD. Since θ cannot exceed a half period, the binaural system may safely assume that the correct delay is the one described by the smallest difference between the apparently leading and lagging waveforms. However, for frequencies at and beyond the point at which the ITD equals a half period (about 770 Hz for sources on the aural axis), both the smaller and the larger values of θ are possible solutions, which makes localization based on the ongoing ITD ambiguous. For the frequency shown in Fig. 5*b*, $\theta = \pi$, and so while the magnitude of the azimuth is unique, there is no way to tell from the ongoing waveforms whether the displacement is to the left or to the right. With still higher frequencies such as in Fig. 5*c*, the number of possible solutions is even greater, for not only is it impossible to know which side is leading or lagging, but also the number of possible θ's is inflated as more peaks in the waveform fall within the range set by the maximum possible ITD. A different kind of problem affects the usefulness of ILDs as a universal cue for direction. As shown in Fig. 4(*b*), the magnitude of the head's acoustic shadow, and hence the ILD, declines with frequency, suggesting a limited role for this cue at low frequencies.

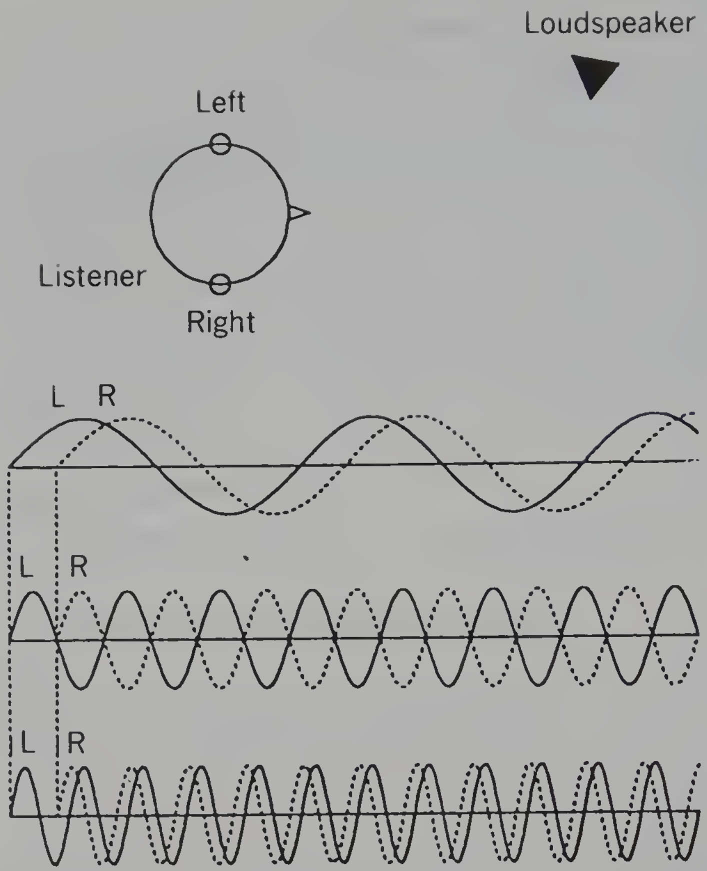


Fig. 5 Schematic representation of ambiguity in the ITD introduced by raising the frequency of a tonal signal presented from a single source.

The traditional solution offered to allay these constraints is called the "duplex theory" of localization.[18] According to this theory, localization is based on two separate mechanisms, one sensitive to ITDs in low frequencies and one to ILDs in high frequencies. Often cited in support of this theory is the accuracy of localization for signals presented from various azimuths in the horizontal plane;[19,20] these data show that the best performance is with tones below 1 kHz and above 3 kHz. Additional support for the low-frequency part of the theory is found when listeners detect an ITD in a continuous tone presented through earphones (see Fig. 6), an ability that disappears for signals above about 1200 Hz. However, a similar argument cannot be made in support of the notion that the ILD is a high-frequency cue, since the ability to detect an ILD presented through earphones shows little or no effect of frequency.

The longevity of the classic duplex theory partly reflects a tradition of research using very narrow-band signals. However, more recent studies have shown the theory to be inadequate for listening with wider bandwidths. One reason is that while an ITD can lead to phase ambiguity with a tonal stimulus, that problem can be resolved through comparisons of interaural time across frequencies (see Section 9.1) or between peaks in the stimulus envelope (see Section 4.3). Also, it is clear

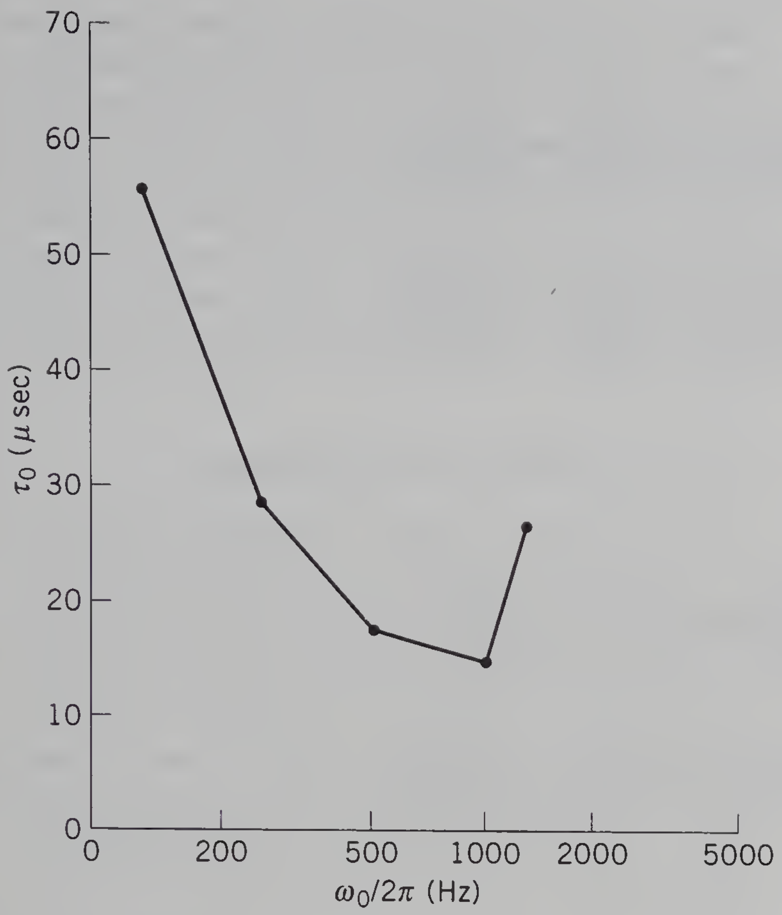


Fig. 6 Jnds for ITD for pure tones presented through earphones (data from several studies compiled in Ref. 2).

that the binaural system sometimes places special significance on timing in the signal's onset, an emphasis so strong that it can preclude the use of ITDs in the ongoing portion of the waveform (see Section 6). Finally, while the duplex theory stresses separate importance of cues in the low and high spectral regions, studies with broadband stimuli have shown a "dominance region" for ITDs in the vicinity of 600 Hz[21] as well as an asymmetry whereby ITDs in the low frequencies dominate localization, regardless of interaural differences in the high frequencies.[4,22]

3.2 Stimulus Onsets

Clearly, binaural information in the signal's onset has special significance for localization. This is easily shown by the case in which a signal is produced from a single source in a reverberant environment. Then, the received stimulus is the sum of multiple wavefronts, one reaching the ears by the most direct pathway and others arriving along secondary or reflected pathways. Despite this multiplicity of binaural information, the listener usually perceives a sound distinctly localized at the position of the source. This domination of localization by the stimulus onset, usually called *binaural precendence*, is discussed in more detail below (see Section 6).

3.3 Sensitivity to Elevation

The most important frequencies for discriminating elevation are in the region above 4–8 kHz, where reflections of the pinna are important, although reflections from the head and torso can also affect frequencies below 2 kHz.[15] To the extent that listeners know how elevation affects the so-called head-related transfer functions (HRTFs), they can localize vertically in the median plane on the basis of purely *monaural cues* by comparing the spectrum of the actual source to that of the received signal.[23] Such comparisons may also be used to perceive azimuth, an especially important ability for listeners with unilateral deafness.[24] Evidence that listeners use such information was found when microphones were placed in listener's ear canals to record stimuli presented from various elevations in the median plane. When the sounds were played back through earphones, subjects were often better at localization if the recordings were from their own ears.[25] However, one must be cautious when interpreting monaural spectral information, because asymmetries of the head and pinnas can produce useful binaural comparisons even with stimuli on the median plane.[26,27]

Use of the pinna transfer functions to discriminate elevation requires comparisons between the relative strengths of components at various frequencies. Thus, it is not surprising that performance is quite poor with narrow bandwidths. For these, the listener may act as though the narrow band is simply a peak in the spectrum of a broadband signal and misperceive its elevation as one for which the HRTF normally accentuates that particular frequency. If so, the sound appears to come from one of these frequency-determined "listening bands" instead of from its actual location.[4,28,29]

3.4 Sensitivity to Distance

The primary effect of increasing the distance to a source is a reduction in the level of the received signal. The sound wave radiating away from a point source forms a sphere whose surface area increases in proportion to the square of the distance traveled. Thus, except in the acoustic near field, the spread of the wavefront over an increasingly larger area results in a loss in level of 6 dB per doubling of distance. In addition, there is a dispersive loss due to energy absorption by the medium. This loss increases with the square of frequency, thus acting as a distance-related low-pass filter. Still more low-pass filtering may be found in cluttered environments such as forests, where objects larger than the wavelength of a sound cast an acoustic shadow.[30]

In an acoustic free field, that is, in an open field in which there are no reflective surfaces, the frequency-dependent changes in level from source to receiver constitute the only information about distance. In a reverberant acoustic environment, the listener receives sounds along secondary pathways from reflective surfaces. Differences between signals in the direct and secondary pathways also carry information about distance. Because their respective levels are related by the ratios of the path lengths while their respective times of arrival are determined by differences between the path lengths, source distance can be found by solving the simultaneous equations for level and time. It has been argued[31,32] that a single variable, the ratio between direct and reflected sound pressure levels, is the primary cue for distance. Architectural acousticians have a special interest in the perceptual effect of reverberation because of its importance for the sense of acoustic spaciousness.[4]

Just as knowledge of the spectral effects introduced by one's head, pinnas, and torso is important for discriminating elevation, it is necessary to have knowledge of the environment in order to be an effective judge of distance. First, knowing the spectral characteristics and overall level of a stimulus at its source provides absolute information needed for comparison, much as a visual observer's knowledge of the size of an object is needed to interpret the retinal image. Second, optimal use of reverberation requires that the listener know the absorptive/reflective characteristics of the surfaces in the environment. Perhaps such knowledge can be obtained from experience, but little is known on that topic.

4 BINAURAL SENSITIVITY MEASURED USING EARPHONES

4.1 Externalization

Traditionally, studies of binaural processing have used stimuli presented via earphones to gain independent control of ITDs and ILDs, something that cannot be done in the open field. A condition in which a stimulus is presented through a single earphone is called *monotic*, whereas a condition in which identical stimuli are presented through both earphones is called *diotic*. Conditions in which different stimuli are presented to the two earphones are called *dichotic*.

When signals are presented to the two ears, if they are sufficiently similar in frequency and if the difference in ITD is not too great, they are perceived as a single entity or a *fused auditory image*. When heard from an external source, these images are typically externalized, that is, reported to be from sources "outside of the head." On the other hand, images produced via earphones are typically reported to be "inside of the head" (although special cases are discussed below in Section 7 for earphone-generated signals that include the transfer functions of the listener's pinna). For this reason, different words have been used for the two cases: Sounds heard as being outside the head are said to be *localized*, whereas those heard as being inside the head are *lateralized*. Current research shows that (1) shaping of the stimulus spectrum to match the effects of the head and pinnas, (2) dynamic recalculation of the interaural cues in response to head movements, and (3) inclusion of room acoustics can all help to make the sounds seem more natural and outside of the head.

4.2 Discrimination versus Salience

When discussing the relative importance of interaural cues, a distinction should be made between the *discriminability* of the cue and its *salience*. The first refers to the how well a *change* in ITD or ILD is detected. It is generally measured as a *just-noticeable difference* (*jnd*) in the cue of interest. Salience refers to the effect of a binaural cue on perceived location, that is, the amount of lateral displacement of the image produced by an interaural difference. Generally it is measured by asking listeners to indicate where they hear an image with an acoustic or visual pointer.

4.3 Interaural Time Difference

Discrimination with Pure Tones Figure 6 shows a sample of jnds found with pure tones presented through earphones (Refs. 33 and 34; by Ref. 2). Human subjects are extremely sensitive to such differences in low frequencies, requiring changes in ITDs of only a few microseconds for detection. However, such accuracy generally requires considerable training in the task. Consistent with the duplex theory, this task is impossible for tones above about 1200 Hz. Interestingly, for frequencies below about 1000 Hz, the threshold declines with frequency. While most likely due to greater variance in the neurotemporal code at low frequencies, it may also reflect accommodation to the physical enhancement of the cue due to creeping waves (see Section 2.1). Interaural time difference discrimination increases with stimulus level,[35] most likely because more auditory neurons are entrained to the stimulus,[36] allowing transmission of more information for binaural comparison.

Sensitivity to Cues in the Envelope While the duplex theory's restriction on the use of ITD clearly holds for continuous, unmodulated pure tones, that is not typically the case for spectrally complex signals. Rather, ITDs are readily detected in high-frequency waveforms as long as there are low-frequency variations in the stimulus envelope.[37] Figure 7 describes one such signal, a sinusoidally amplitude modulated (SAM) tone. In this

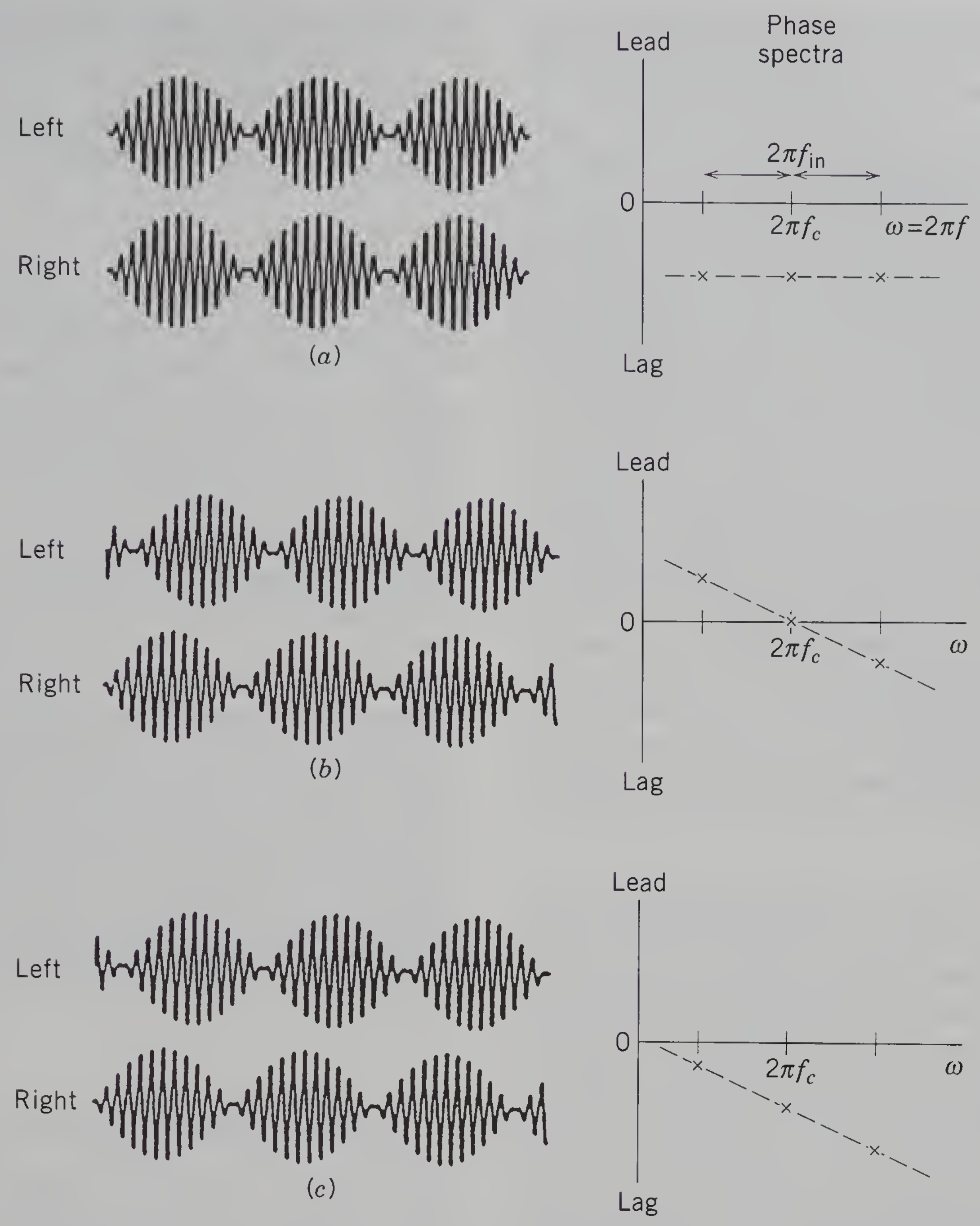


Fig. 7 Waveforms and interaural phase spectra of SAM tones presented dichotically with interaural delays of (*a*) the carrier alone, (*b*) the envelope alone, or (*c*) the whole waveform. (Reprinted from Ref. 43.)

case, a single, high *carrier* frequency (f_c) is modulated at a low frequency (f_m); SAM results from multiplying $\sin(2\pi f_c t)$ by $[1 + \sin(2\pi f_m t)]$. The fine structure of the resulting waveform varies at the rate of f_c while its amplitude varies at the rate of f_m. The envelope, as defined by a line connecting adjacent peaks in the waveform, is a raised sinusoid with frequency f_m—thus the name sinusoidal amplitude modulation. The frequency spectrum of this waveform consists of a component at f_c plus two side bands of half its amplitude at $f_c - f_m$ and $f_c + f_m$. In a closely related stimulus also tested for lateralization, the carrier is suppressed by omission of the constant 1 in the modulator, leaving only the two side bands.[38] In either case, the listeners can detect an ITD in the envelope of a modulated tone, even when all of the resulting frequencies are above the usual dividing line of the duplex theory (e.g., Refs. 38–42). The right side of Fig. 6 shows the separate effects on the interaural phase spectrum of SAM produced by delaying just the f_c, just the f_m, or the entire waveform.[43] For high f_c's, an ITD in just the carrier cannot be detected.[39]

The ITDs can also be detected in bands of high-frequency noise (e.g., Refs. 33, 38, and 44) and single and repeated high-frequency clicks (e.g., Ref. 45–47), with jnds rivaling those found with low-frequency pure tones. However, caution must be applied when designing such experiments because it has also been shown that listeners are able to utilize ITDs in the highly attenuated low-frequency skirts of nominally high-frequency signals.[44,46]

Resolution of ITD across Binaural Range In vision, spatial discrimination is most accurate in the central field of view, showing a decline in the density of retinal innervation from fovea to periphery. Similarly, localization of tones presented in the free field[20] is most accurate near the median plane. However, studies using earphones to study sensitivity to changes in ITD in both tones[35] and clicks[48] have found that the jnd increases only a small amount in response to large changes in the baseline ITD. This suggests that different mechanisms are responsible for center weighting in the two sensory modalities. Unlike the case in vision, where an angular change in the direction of a source produces a comparable shift of the retinal image, lateral displacement of a sound in the free field produces an ITD that is a function of the cosine of the displacement.[11] Thus, an angular shift of a source near the aural axis has a much smaller effect on ITD that a comparable shift of one near the median plane. This means that the apparent center weighting of ITD sensitivity in the free field is based primarily on the physics of sound.

Binaural Beats When two slightly mistuned low-frequency tones are presented one to each ear, they are perceived as a single tone that moves from one side to the other at the rate of the frequency difference. Because of their similarity to amplitude beats heard due to acoustic interference when tones are presented to a single ear, these are called *binaural beats*. As the difference frequency is increased above about 4 Hz, the beats turn to a "roughness" that itself disappears at differences of about 10 Hz.[49,50] This stimulus has been used extensively in physiological studies of the binaural nervous system[51,52] to assess the ITD sensitivity of individual neurons. Binaural beats can also be conveyed by an interaural difference in the envelope frequencies of an amplitude-modulated high-frequency carrier; interestingly, this can occur even with disparate carrier frequencies.[53]

4.4 Interaural Sound Pressure Level Difference

Discrimination The ILD is defined as the difference (in decibels) between the sound pressure levels at the two ears. Typically, jnds are found by presenting +ILD/2 dB to one ear and −ILD/2 dB to the other in an attempt to maintain constancy of overall loudness. A variety of experiments using tones, clicks, and noises have shown the jnd to be on the order of 1 dB, regardless of frequency.[35,54] This value holds even when the stimuli are independent noises presented to the two ears.[55]

Resolution of ILD across Binaural Range Similar to the case with ITD, there is little effect of the baseline ILD on the jnd with tones[35] or clicks.[56] Thus, binaural resolution seems reasonably evenly distributed over a large range of interaural differences.

4.5 Salience

Methods used to measure the salience of an interaural difference ask listeners to indicate the lateral displacement of an auditory image. Techniques range from paper-and-pencil drawings of the image location (e.g., Refs. 57 and 58) to direct matching of the location of the image with an acoustic "pointer" whose own lateralization is set by an ITD[59] or an ILD[60,61]

Figure 8 shows the effect of ITD on the perceived lateral displacement of low-frequency tones. These data demonstrate that the relation between ITD and lateral displacement is reasonably linear until the delay approaches a half period of the tone, at which point most listeners report two images, one at each ear. Salience due to ITD has been shown to diminish above 1200 Hz as predicted by the duplex theory.[62] The salience of ILD is shown in Fig. 9 (composite from Ref. 63). Its effect on image displacement is fairly linear up to about 10 dB. Given the strong effect of frequency on the magnitude of the ILD in nature, it is intriguing that the salience of ILD in lateralization is essentially independent of frequency.[58]

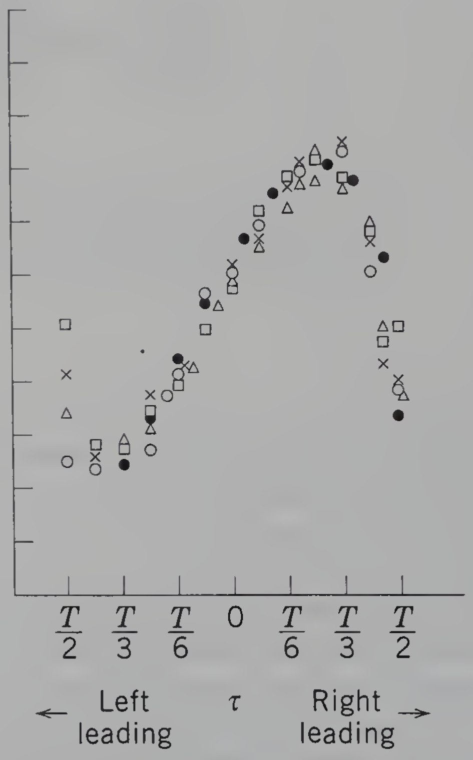


Fig. 8 Salience of ITDs in tones defined by listeners' drawings of the intracranial locations of the images of tones ranging from 200 to 1200 Hz plotted as a function of the tone's period T. (Data from Ref. 57 as adapted by Ref. 2.)

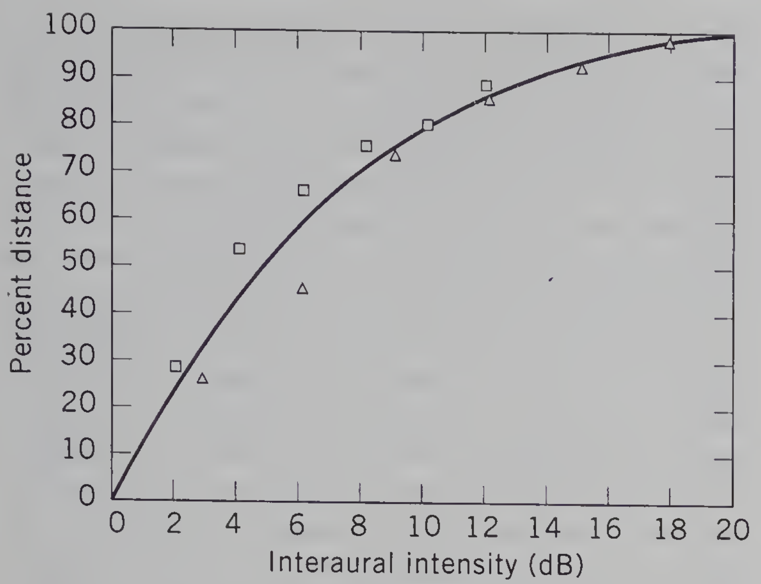


Fig. 9 Salience of ILD in tones defined by listeners' drawings of the intracranial locations of the images. (Data are a composite from Refs. 58 and 145 as redrawn in Ref. 63.)

Studies using earphones to investigate the extent of the lateral shift of an auditory image due to ITDs in high-frequency bands of noise or envelopes of high-frequency sinusoidal amplitude modulation (e.g., Refs. 4 and 64) have found that ITDs have much less effect on apparent displacement than might be predicted from their effectiveness for discrimination (Section 4.3). This observation underlines the importance of considering both measures of performance when assessing the potency of interaural cues.

4.6 Interrelations between Binaural Cues

Time–Intensity Trading Independent control of ITD and ILD with earphones allows for direct comparisons of the relative effectiveness of these cues in what is usually called the *binaural time–intensity trading ratio.* Earphones allow one to present values of the ITD and ILD that are either *consonant*, with both cues favoring the same side, or *dissonant*, with the ITD favoring one side while the ILD favors the other. The trading ratio is commonly reported in linear units of interaural time and the logarithm of the interaural sound pressure ratio (microseconds per decibel). The trading ratio has been measured in many ways. These include *image-centering* tasks (e.g., Ref. 65) in which a listener adjusts one of the interaural cues in order to overcome the effect of a dissonant value of the other, *pointing* tasks (e.g., Ref. 66) in which the listener adjusts a second sound until the lateral position of this "pointer" matches that of the signal, *comparison* tasks[60] in which the trading ratio is inferred from values of the ITD and ILD that produce the same level of discrimination, and *detection* tasks[67] in which the listener discriminates between a diotic standard and a combination of dissonant ITDs and ILDs whose values are chosen to produce an optimal trade. In the latter case, trading ratios are inferred from the combination that produces poorest performance. A strong interpretation of time–intensity trading says that at some point in the auditory system ITD and ILD are neurally equivalent. However, the demonstration[67] that no combination of the two cues was completely indiscriminable from the diotic shows that the strong interpretation is incorrect. Typical trading ratios are on the order of 20–50 μs/dB for both low-frequency tones and clicks. They tend to be larger with high-frequency stimuli, especially at low levels.

Double Images An interesting result from some of the trading-ratio experiments is that listeners sometimes say that they hear two distinct images, each described by its own trading ratio.[59,65,66] In such cases, the second image has a trading ratio close to 0 μs/dB, showing virtually no effect of ILD. Historically, this image has been called the "time" image to differentiate it from the "intensity" image that is more strongly affected by ILD. These results are in line with the view that there is not a single neural code that includes both ITD and ILD.

4.7 Cross-Frequency Interference

In everyday life, we are often confronted with multiple stimuli, each with its own spectrum and location. Localization of a source in such an environment requires that the listener be able to separate the binaural information arising from each of the potential sources.

Interference with Binaural Discrimination Experiments designed to study interference have examined the discrimination of ITD in a "target" stimulus in the presence of other, "interfering" stimuli. Results show that low frequencies tend to dominate high; for example, a narrow band of diotic noise centered on 500 Hz reduces discrimination of the ITD in a band of noise centered on 4 kHz while the reverse is not true.[38] Cross-frequency binaural interference is generally greatest when both target and interferer are turned on simultaneously, although interference can occur even when the onset of the interfering stimulus precedes that of the target by as much as a quarter of a second (e.g., Refs. 68 and 69). However, the effect is greatly reduced if the interferer is continuous.[70] Interestingly, despite the interference found with onset asynchrony, listeners report hearing the targets and interferers as separate acoustic events.[68] Thus, the sounds are "segregated" according to their subjective appearance but not according to their binaural interaction.

Other Stimulus Factors Many important stimuli such as voices and the sounds of musical instruments have harmonic spectra. Because of this, interest has arisen in the harmonic relationship between targets and potential interferers. For example, one report[71] found interference when their tonal targets and interferers were both at low harmonic numbers of a common fundamental, but not when the tones were inharmonic. Another[72] found that a harmonically related interfering tone reduced the binaural *masking-level difference* (see Section 5 below) for a low-frequency tonal signal but did not extend this to inharmonic interferers. However, others (e.g., Ref. 73) have reported lateralization interference even from inharmonic tones. A potential explanation for these discrepancies lies in the considerable individual differences between subjects with respect to the effects of harmonicity.[72] Perhaps there are multiple listening strategies that determine the nature and amount of binaural interference.

5 BINAURAL MASKING

Binaural masking refers to a set of special conditions in which signal detection is dependent on interaural differences created when the signal is added to the masker. For an example, consider the case schematized in Fig. 10. In the leftmost drawing, the signal is a 500-Hz tone burst presented diotically through earphones while the masker is a continuous wide-band noise, also diotic. A trial consists of a pair of intervals with the signal presented in only one. The listener's task is to detect which interval contains the signal. By convention, this condition is referred to as **NoSo**, with the small O's indicating that there are no interaural differences in the noise or signal. The signal's level is set to produce 75% correct detections. The center drawing depicts a *binaural condition*. Here, removal of the signal from the right earphone has produced the case called **NoSm**, with the small M indicating that the signal is monotic. Note that the signal-to-noise ratio in the left earphone is unchanged from **NoSo** but now there is additional information in the form of interaural differences between signal plus noise at the right ear and noise alone at the left. Based on this interaural information, detection improves to the point that the signal must be *attenuated* by 9 dB if the listener is to achieve only 75% correct detections. This reduction in signal needed for detection in the binaural condition is referred to as a *masking-level difference (MLD)*, thus, one would say that the MLD for **NoSm** is 9 dB. Finally, in the rightmost drawing, the noise is also removed from the right earphone. This condition is called **NmSm**. It offers no interaural information for detection, which means that the signal level must be restored to that used under **NoSo** for a performance of 75% correct. Excellent treatises on this subject include Refs. 1, 2, 74, and 75.

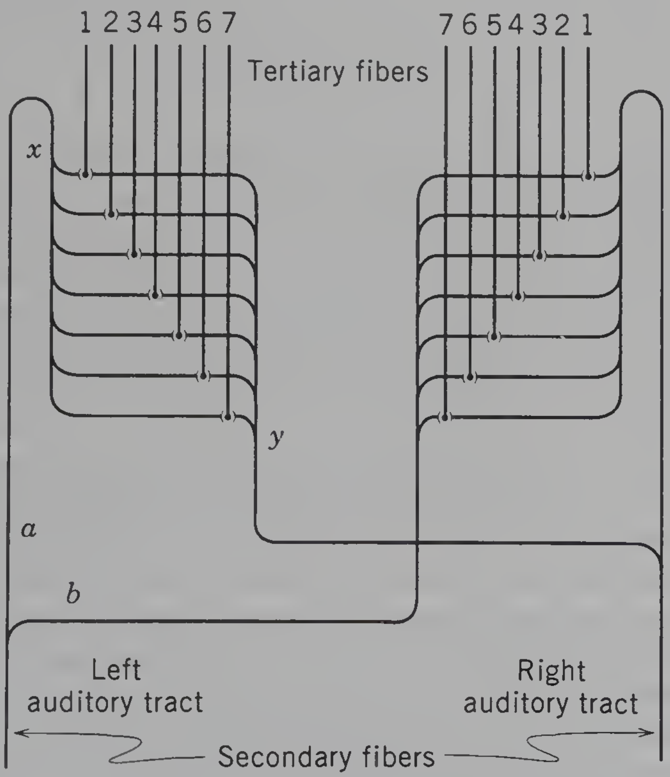


Fig. 10 Schematic drawing of a hypothetical neural mechanism proposed for computing a binaural cross correlation. (Reprinted from Ref. 118.)

5.1 Binaural Conditions

The traditional naming convention[74] for the binaural conditions takes the form **MxSy**, with x and y referring to interaural relations in the masker (**M**) and signal (**S**), respectively. (As noted above, for a noise masker, **N** is substituted for **M**.) When there are no interaural differences, x or y are indicated by the modifier o. The reference condition against which binaural conditions are compared in computation of the MLD is uusually **NoSo**, but the purely monaural condition, **NmSm**, serves equally well. Most commonly tested of the binaural conditions is **NoS**π, in which a signal reversed in phase interaurally (π radians) is presented in a diotic noise. Other conditions that produce MLDs include reversing the noise interaurally but not the signal (**N**π**So**) and reducing the interaural correlation of the noise by mixing interaurally independent noise with the diotic masker. A totally uncorrelated masker is indicated by a u as in **NuSo**.

5.2 Masking-Level Difference

Several conclusions can be drawn from the literature on binaural masking. With a diotic masker, the MLD increases as the interaural phase of the signal goes from

O to π. When the signal is a low-frequency tone presented in wide-band noise, the MLD for **N**O**S**π is about 15 dB. Various other combinations of interaural phase and level differences in the signal produce MLDs within the range bounded by zero and the MLD for **N**O**S**π (see Ref. 2). Interaurally reversing the masker instead of the signal (**N**π**S**O) produces an MLD of about 12 dB.[74] Reducing the correlation of the noise from +1, as it is in **N**O**S**π, or from −1, as it is in **N**π**S**O, reduces the MLD. With a correlation of zero, as in **N**U, the MLD is about 3–4 dB.[76–78] When the masker is a broadband noise, largest MLDs occur for signal frequencies around 400–500 Hz. The MLD decreases with frequency up to around 1500 Hz where the MLD for **N**O**S**π reaches an asymptotic value of about 3 dB.[79] Reducing the bandwidth of the masker leads to much larger MLDs.[80,81] The MLD is little affected by the signal's duration.[82,83]

If both the signal and the masker are tones, the magnitude of the MLD depends on the signal-to-masker phase as well as the interaural phase.[84] Interestingly, the MLD for speech reception depends upon the requirements of the task. For simply detecting the presence of speech in noise, the data are predictable from experiments with tonal signals and dominated by the lower frequencies. However, when the requirement is to comprehend the message, the MLD is much smaller (on the order of 6 dB), probably because so much of the information necessary for discrimination is carried in the higher frequencies, which show less release from masking.[85,86]

6 ONSET EFFECTS AND BINAURAL PRECEDENCE

One of the oldest topics in the field of binaural hearing has been studied under several names: *binaural precedence effect*, *Haas effect*, and *law of the first wavefront*. These terms refer to the listener's subjective report that sources seem to be located according to the azimuth of the first wavefront despite delayed versions of the signal arriving along secondary pathways. Excellent reviews includ Refs. 4, 87, and 88.

6.1 Precedence

In the best known paradigm (e.g., Ref. 89), a pair of dichotic clicks is presented through earphones, with the second click following the first by a short interval. The lateralization of the *single image* produced by perceptual fusion of both pairs of clicks is much more affected by the ITD in the first click than in the second. Other demonstrations of special importance given to onsets include comparisons between the effects of ITDs in the onset and the ongoing portions of a burst of noise[90–92] and comparisons between the detectability of ITDs placed at various temporal locations in a gated noise.[88]

6.2 Fusion of a Primary and Secondary Source

Dramatically different effects have been described for three different regions of interstimulus delay.[4] Delays of up to about 1 ms produce what Blauert[4] calls "summating localization," with the primary and secondary stimuli combining to produce a single image lateralized at an azimuth based on the direction of both sources. The second region describes classical precedence, with the signal seeming to come from the azimuth of the primary source. The extent of this region depends upon the nature of the signal, extending to tens of milliseconds for long-duration, complex sounds such as speech and music. He calls the upper boundary of this region the "echo threshold" to indicate that still longer delays produce two distinct sounds heard as a source and echo. Important for everyday listening are demonstrations that onset transients foster accurate localization in rooms.[93] Those authors postulate that transients "trigger" the precedence effect, which in turn diminishes interference from reflections.

Although listeners may report hearing only a single source under conditions of precedence, they are not entirely unaware of the secondary pathway. This can be demonstrated in one's living room by sitting close to one of the loudspeakers of a stereo system while listening to music. Based on precedence, all of the sound seems to come from the nearer speaker. However, if the second speaker is disconnected, the changes in loudness and even timbre are easily noted. In laboratory settings, subjects are able to tell whether a secondary pathway is on the right or left of the primary,[94] and listeners asked to point toward the direction of a transient source in the presence of a reflective surface show biases of the image toward the direction of the echo.[95]

6.3 Postonset Lateralization and Binaural Adaptation

The relative importance of each portion of a signal with respect to lateralization can be found by comparing the magnitudes of the interaural differences needed to produce comparable performance with a number of different durations. Equivalence of information across portions means that the variance of the limiting noise declines in proportion to duration, thus decreasing the interaural difference needed for constant detectability according to the square root of duration. This argument has been applied to studies of the effectiveness of ITDs and ILDs in trains of low-frequency clicks,[96] noise,[97] low-frequency tones,[98] amplitude-modulated sinusoids,[41,42] and trains of high-frequency clicks.[47,99] For low tonal frequencies, low modulation frequencies, or low repetition rates of clicks (below about 85 Hz), each portion of the signal is equally effective. However, for higher frequencies or rates, the relative effectiveness of each portion declines beyond the stimulus onset. This decline has been called *binaural adaptation*.[100] It takes the form of a compressive power function whose exponent declines with increases in the frequency or rate. When the latter exceeds several hundred hertz, the exponent goes to zero, making the response essentially that of an onset detector, that is, unaffected by ITDs or ILDs in the ongoing portions.

An interesting problem related to binaural adaptation is the question of when the system recovers from adaptation and resamples the spatial environment. Studies indicate that, in addition to termination of the stimulus, other auditory "triggers" can induce recovery from adaptation. These include short gaps in the stimulus, short bursts of noise, and brief tones added to the stimulus.[101] One interpretation is that the trigger is heard as a new stimulus requiring localization. This view implies the presence of a "top-down" process that discriminates between sounds that are a part of old (ongoing) stimuli and those that are new. It is generally felt that a process of this kind must also be a part of precedence since the binaural system must segregate primary and secondary acoustic pathways in order to reduce the importance of the latter.[100] The idea that higher order perceptual strategies sort out potentially conflicting binaural cues is akin to what has been described as the *plausibility hypothesis*[93] and is prominent in many views of precedence (e.g., Ref. 102). Recent work with artificially created secondary pathways supports this position[103] by showing that an unexpected transposition of the hypothetical source with its reflection produces a period of perceptual confusion during which the two are heard as distinct and separate sources.

7 VIRTUAL ENVIRONMENTS

Modern computing techniques have led to a new way of examining perceptual processes. In what are called *virtual environments*, artificially generated stimuli are used to simulate a real environment (for extended discussion see Refs. 104 and 105). This technology has fostered widespread interest in applications dedicated to auditory entertainment,[106] including creation of electronic devices that create the sense of being in a concert hall and the illusion of localization and spaciousness in films and computer-based games. Other applications for virtual technology include solutions to the so-called tele-operator problem wherein sensors in an unmanned environment supply information to a remote human operator.

Virtual environments have also been used in basic research in hearing. These applications include simulation of a virtual source with simultaneous presentations from multiple loudspeakers,[107] simulation of auditory movement,[108] and localization with artificial echoes.[109] However, the major thrust has been toward creation of sounds through earphones that approximate what is heard in an open field. Accordingly, stimuli are filtered through HRTFs (see Section 3.3) meant to reproduce the effects of reflections and resonances by the shoulders, head, and external ears that would accompany stimulation in the open field. Where properly applied, signals so filtered produce apparent localizations in azimuth and elevation quite similar to those produced in a free-field task.[4,22,110,111] While there is a sense of externalization with such signals, there is often a reduced sense of stimulus distance, with sounds appearing to be relatively close to the head. One hypothesis about this problem stresses the importance of sensitivity to dynamic changes in the interaural cues produced by movements of the head. By using a sensor to measure the position of the listener's head and updating the HRTF accordingly, it is possible to greatly enhance the naturalness of virtual environments heard under earphones.[112] Difficulties encountered so far in externalizing the auditory images heard through earphones suggest that we must also take account of the fact that everyday listening is done in reverberant environments. It appears that for virtual stimuli to be perceived as being appropriately distant, the recordings should be obtained in a reverberant environment.[112,113]

8 MOTION

Auditory motion can be thought of as a time-variant change in the apparent location of a source. The methods used for such studies include creation of artificial motion induced through earphones by variation of the interaural parameters,[114] real motion through movement of loudspeakers in space,[115] and artificial motion created in space by changing the phases of sinusoids from a pair of speakers.[108] In general, the results show that the auditory system is relatively slow in detecting changes in the binaural parameters, a state referred to as binaural "sluggishness."[116] Detection of the direction of motion is an inverted U-shaped function of velocity, peaking at about 2°/s.[117] Sluggishness is often thought to represent persistence in the mechanisms that extract interaural disparities, as though the ITDs and ILDs are smoothed or integrated via separate low-pass filters.[118]

9 THEORIES OF BINAURAL PROCESSING

Although the ability to detect ITDs in amplitude-modulated high frequencies disproves a strong version of the duplex theory, the theory nevertheless provides a good first-order description of the majority of binaural effects, especially with narrow-band sounds. That is, ITDs are the effective cues in low-frequency sounds and ILDs in high-frequency ones. Comprehensive reviews of theories of binaural hearing include Refs. 2 and 4–6.

9.1 Cross-Correlation and Interaural Time Difference

According to most theories of localization, the central nervous system encodes the ITD by computing a binaural cross-correlation function. Correlation requires integration of the products of the inputs from the two ears while generation of the entire function requires separate calculations for each interaural delay. A neural mechanism first proposed for this purpose by Jeffress[119] is depicted in Fig. 10. Multiplications are done by synaptic coincidence detectors that generate neural outputs in response in simultaneous activity in pathways from the two ears. The cross correlations are then integrated separately for each value of the ITD. This is done by accumulating coincidences at the junctions of different pairs of channels from the two ears, each with its own sets of delays as depicted by differing lengths of the hypothetical neural pathways. As shown, the model produces the cross-correlation function for a single frequency, but by assuming that other frequencies are served in parallel, it is possible to visualize a spatially coded neural array in which cross correlations are plotted as a function of frequency. In the nearly half century since this model was proposed, physiological measures made in mammals and birds have found cells in the central auditory pathways whose rate of response depends upon characteristic ITDs,[51,52,120,121] and recently anatomists have reported neural arrays in the brains of birds that are quite similar to Jeffress's[119] model.[122,123] Additional support at the level of behavioral analysis can be seen in the success of cross-correlation models in predicting (1) the binaural release from masking;[124] (2) discrimination and salience of interaural delay;[125] (3) detection of changes of interaural correlation;[126,127] (4) failure of the precedence effect for low-frequency, narrow-band signals;[128] and (5) the compactness/diffuseness of binaural images produced in various environments including concert halls.[5]

Ambiguity A shortcoming of simple cross-correlational models is their lack of a provision for the phase ambiguities encountered with narrow-band signals. A good example of this problem is seen with a pure-tone stimulus for which the cross-correlation function repeats at values of the ITD equal to the period of the tone. The traditional assumption has been that such ambiguity is defeated by emphasizing the smallest of the potential ITDs, this done by multiplication of the cross-correlation function by a weighting function centered on ITD = 0.[124] Thus, a 50-μs ITD favoring the right ear (in a 1000-Hz tone with period 1000 μs) is given more weight than the next most central peak at 950 μs favoring the left. While center weighting is important, interest has also focused on the suggestion that ambiguity is resolved by finding the value of the ITD that is common to all of the signal's frequencies.[75] An example of results from a computational model that demonstrates this notion[129] is shown in Fig. 11. At each frequency, ambiguous peaks in the correlation function radiate from the true ITD at intervals of the period of that frequency. However, the peaks are aligned across frequency at the single value of the true ITD shown by the vertical straight line. Compelling demonstrations of the importance of such "straightness" have recently appeared[129] in behavioral studies in which specially tailored signals place center weighting in opposition to straightness. Electrophysiological studies of the midbrain of the barn owl have found individual cells tuned to specific ITDs, regardless of frequency.[130] Tracing backward to a more peripheral stage of the midbrain has then shown that each ITD cell represents convergence from an array of neurons whose combinations of best frequency and best interaural phase described the appropriate linear relation between frequency and phase,[131] that is, straightness in the cross-correlation function. Together, these observations offer strong support for a model of human lateralization that incorporates a second level of coincidence detection across frequency.[132]

Lateral Inhibition and Cross Correlation An interesting version of a binaural cross-correlation model[122,133] utilizes a process known to be of great

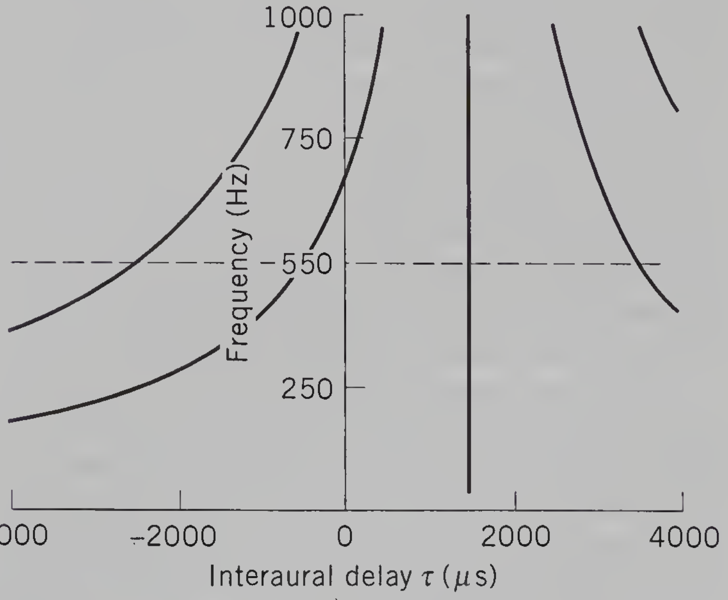


Fig. 11 Peaks in the binaural cross correlation plotted as a function of frequency (Stern et al.,[129]). The stimulus is a wide-band noise interaurally delayed by an ITD of 1500 μs.

importance in visual processing, the *lateral inhibition* of some neural activity by others. In this case, the outputs of coincidence detectors, *a la* Jeffress,[119] inhibit the inputs for other detectors, thus increasing their own importance. This approach also stresses the importance of separate monaural pathways acting in conjunction with the binaural. The power of this model[133,134] is in its ability to deal not only with a wide variety of binaural phenomena including the interaction between ITD and ILD but also with binaural precedence, for which it suggests a simple, relatively peripheral mechanism.

Interaural Sound Pressure Level Difference Models of the mechanisms for coding ILD are generally simpler than those for ITD. Often they simply assert that the location of the localized image is weighted by the difference in level. One suggestion[135] is that the lateral displacement of an image reflects the number of neurons in the central pathways tuned toward one side or the other of the midline, with excitation (E) from one ear being opposed by inhibition (I) from the other. In this case, either ITD *or* ILD could affect the number of neurons tuned toward the side of the leading or more intense signal. As seen in Chapter 111, excitatory–inhibitory (EI) interactions have been identified in the binaural pathways. Because the cross correlation itself is not affected by the levels of the stimuli in each ear, such models have had to incorporate an additional mechanism such as a level-dependent weighting function to account for the effects of ILD.[128,136]

9.2 Masking-Level Differences

Historically, the issue of what is detected in binaural masking has been approached from two general perspectives. In one (e.g., Ref. 74), detection in conditions that produce an MLD occurs when the signal produces a difference in lateralization. Sometimes called "vector" theories, such models assume that noise is made narrow band by the cochlear filters and so can be treated as a slowly changing sinusoid. The models then compute ITDs and ILDs from vectorial summation of the signal and masker.[74,84,137] The idea is that the ITDs, and perhaps the ILDs, act in consort, producing patterns in the neural representation of binaural space that are different for the masker alone and for the signal plus masker.

In different perspective Durlach[138] has argued for a more quantitative definition of the internal noise that limits binaural detection in his *equalization–cancellation* (*EC*) theory. This model proposes neural transformations that "equalize" the maskers from the two ears and then take the difference, leaving as a residual any signal not equalized by the first procedure. An important feature of the EC model is that incorporation of the internal noise, in the form of "jitter" in both the amplitude and time domains, defines the factors that ultimately limit performance and allows the model to be tested over a host of experimental parameters.[139] It can be shown that the effective stimulus in each of the two models is essentially the same.[140]

10 FUTURE DEVELOPMENTS

One of the most appealing features of sound localization for theorists over the past century has been the opportunity to speculate about perceptual processes that must, because of their reliance on inputs from the two ears, take place in the central nervous system. Examples of the great variety of physiological models currently being proposed to merge with psychoacoustic findings include neural feedback loops and their potential for understanding onset effects,[141] coding of the spectral notch in the pinna transfer function,[142] potential interactions between interaural delay and the cochlear traveling wave,[143,144] and neural bases for cross correlation.[51,52,122] Nevertheless, much still remains to be learned about binaural interaction at higher levels of perception, especially about localization in complex acoustic fields and perception of separable auditory events in a multievent environment. Musicians will want to know more about the role of timbre and localization in determining the quality of a performance just as acoustical architects and engineers will continue to use such data to build better halls.[5] Another issue of concern will be the interaction between spatial vision and spatial hearing as well as other interactions between the basics of audition and such higher order processes as selective attention. Finally, fast computers are opening whole new opportunities for building truly comprehensive theories of auditory function, and models of binaural interactions will be a necessary part of those efforts.

Acknowledgments

This chapter was supported by two research grants from the National Institute for Deafness and Other Communication Disorders (DC 00087 for E. R. H. and DC 000234 for C. T.) and the Air Force Office of Scientific Research (AFOSR 89-0030). The authors thank Eric Jensen, Les Bernstein, Laurie Heller, and Ina Alexandrovsky for their many helpful comments.

REFERENCES

1. H. S. Colburn and N. I. Durlach, "Models of Binaural Interaction," in E. C. Carterette and M. P. Friedman

(Eds.), *Hearing*, Vol IV, *Handbook of Perception*, Academic, New York, 1978.

2. N. I. Durlach and H. S. Colburn, "Binaural Phenomena," in E. C. Carterette and M. P. Friedman (Eds.), *Hearing*, Vol IV, *Handbook of Perception*, Academic, New York, 1978.
3. C. L. Searle, L. D. Braida, M. F. Davis, and H. S. Colburn, "Model for Auditory Localization," *J. Acoust. Soc. Am.*, Vol. 60, 1976, pp. 1164–1175.
4. J. Blauert, *Spatial Hearing*, MIT Press, Cambridge, MA, 1983.
5. Y. Ando, (1985). *Concert Hall Acoustics*, Springer-Verlag, Berlin.
6. W. A. Yost and G. Gourevitch (Eds.), *Directional Hearing*, Springer-Verlag, New York, 1987.
7. H. C. Middlebrook and D. M. Green, "Sound Localization by Human Listeners," *Ann. Rev. Psychol.*, Vol. 42, 1991, pp. 135–159.
8. M. Morimoto and H. Aokata, "Localization Cues of Sound Sources in the Upper Hemisphere," *J. Acoust. Soc. Jpn.* (E), Vol. 5, 1984, pp. 165–173.
9. E. I. Knudsen, "Auditory and Visual Maps of Space in the Optic Tectum of the Owl," *J. Neurosci.*, Vol. 2, 1982, pp. 1177–1194.
10. J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional Sensitivity of Sound-Pressure Levels in the Human Ear Canal," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 89–108.
11. R. S. Woodworth, *Experimental Psychology*, Holt, Rinehart & Winston, New York, 1938.
12. G. F. Kuhn, "Model for the Interaural Time Differences in the Horizontal Plane," *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 157–167.
13. G. F. Kuhn, "Physical Acoustics and Measurements Pertaining to Directional Hearing," in W. A. Yost and G. Gourevitch (Eds.), *Directional Hearing*, Springer-Verlag, New York, 1987.
14. R. K. Clifton, J. Gwiazda, J. A. Bauer, M. G. Clarkson, and R. M. Held, "Growth in Head Size During Infancy: Implications for Sound Localization," *Devel. Psychol.*, Vol. 24, 1988, pp. 477–483.
15. W. E. Feddersen, T. T. Sandel, D. C. Teas, and L. A. Jeffress, "Localization of High-Frequency Tones," *J. Acoust. Soc. Am.*, Vol. 29, 1957, pp. 988–991.
16. E. A. G. Shaw, "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1848–1862.
17. E. A. G. Shaw, "The External Ear," in W. D. Keidel and W. D. Neff (Eds.), *Handbook of Sensory Physiology*, Vol. 1, Springer-Verlag, New York, 1974.
18. Lord Rayleigh (J. W. Strutt), "On Our Perception of Sound Direction," *Philos. Mag.*, Vol. 13, 1907, pp. 214–232.
19. S. S. Stevens and E. B. Newman, "The Localization of Actual Sources of Sound," *Am. J. Psychol.*, Vol. 48, 1936, pp. 297–306.
20. A. W. Mills, "On the Minimum Audible Angle," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 237–246.
21. F. A. Bilsen and J. Raatgever, "Spectral Dominance in Binaural Lateralization," *Acustica*, Vol. 28, 1973, pp. 131–132.
22. F. L. Wightman and D. J. Kistler, "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization," *J. Acoust. Soc. Am.*, Vol. 91, 1992, pp. 1648–1661.
23. J. Hebrank and D. Wright, "Spectral Cues Used in the Localization of Sound Sources on the Median Plane," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1829–1834.
24. W. H. Slatterey and J. C. Middlebrooks, "Monaural Sound Localization: Acute vs. Chronic Unilateral Impairment," *Abst. Assoc. Res. Otolaryngol.*, Vol. 15, 1992, p. 95.
25. R. A. Butler and K. Belendiuk, "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 1264–1269.
26. C. L. Searle, L. D. Braida, D. R. Cuddy, and M. F. Davis, "Binaural Pinna Disparity: Another Auditory Localization Cue," *J. Acoust. Soc. Am.*, Vol. 57, 1975, pp. 448–455.
27. S. R. Oldfield and S. P. A. Parker, "Acuity of Sound Localization: A Topography of Auditory Space. I. Normal Hearing Conditions," *Perception*, Vol. 13, 1984, pp. 581–600.
28. R. A. Butler and N. Planert, "The Influence of Stimulus Bandwidth on Localization of Sound in Space," *Percept. Psychophys.*, Vol. 19, 1976, pp. 103–108.
29. R. A. Butler and C. C. Helwig, "The Spatial Attributes of Stimulus Frequency in the Median Sagittal Plane and Their Role in Sound Localization," *Am. J. Otolaryngol.*, Vol. 4, 1983, pp. 165–173.
30. C. H. Brown and P. M. Waser, "Environmental Influences on the Structure of Primate Vocalizations," in Todt, Goedeking, and Symmes (Eds.), *Primate Vocal Communication*, 1988.
31. G. von Bekesy, *Experiments in Hearing*, McGraw-Hill, New York.
32. D. H. Mershon and J. N. Bowers, "Absolute and Relative Cues for the Auditory Perception of Egocentric Distance," *Perception*, Vol. 8, 1979, pp. 311–322.
33. R. G. Klumpp and H. R. Eady, "Some Measurements of Interaural Time Difference Thresholds," *J. Acoust. Soc. Am.*, Vol. 28, 1956, pp. 311–322.
34. J. Zwislocki and R. S. Feldman, "Just Noticeable Differences in Dichotic Phase," *J. Acoust. Soc. Am.*, Vol. 28, 1956, pp. 860–864.
35. R. M. Hershkowitz and N. I. Durlach, "Interaural Time and Amplitude JND's for a 500-Hz Tone," *J. Acoust. Soc. Am.*, Vol. 46, 1969, pp. 1583–1584.
36. R. Pfeiffer and D. Kim "Cochlear Nerve Fiber Responses:

Distribution along the Cochlear Partition," *J. Acoust Soc. Am.*, Vol. 58, 1975, pp. 1583–1584.

37. D. M. Leakey, B. McA. Sayers, and C. Cherry, "Binaural Fusion of Low- and High-Frequency Sounds," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 222–223.
38. D. McFadden and E. G. Pasanen, "Lateralization at High Frequencies Based on Interaural Time Differences," *J. Acoust Soc. Am.*, Vol. 59, 1976, pp. 634–639.
39. G. B. Henning, "Detectability of Interaural Delay in High-Frequency Complex Waveforms," *J. Acoust Soc. Am.*, Vol. 55, 1974, pp. 84–90.
40. D. McFadden and C. M. Moffitt, "Acoustic Integration for Lateralization at High Frequencies," *J. Acoust Soc. Am.*, Vol. 61, 1977, pp. 1604–1608.
41. J. M. Nuetzel and E. R. Hafter, "Lateralization of Complex Waveforms: Effects of Fine-Structure, Amplitude, and Duration," *J. Acoust Soc. Am.*, Vol. 60, 1976, pp. 1339–1346.
42. J. M. Nuetzel and E. R. Hafter, "Discrimination of Interaural Delays in Complex Waveforms: Spectral Effects," *J. Acoust Soc. Am.*, Vol. 69, 1981, pp. 1112–1118.
43. G. B. Henning, "Lateralization of Complex Waveforms," in G. van den Brink and F. A. Bilsen (Eds.), *Psychophysical, Physiological and Behavioral Studies in Hearing*, Delft University Press, Delft, 1980.
44. L. R. Bernstein and C. Trahiotis, "Detection of Interaural Delay in High Frequency Noise," *J. Acoust Soc. Am.*, Vol. 71, 1982, pp. 147–152.
45. G. G. Harris, "Binaural Interactions of Impulsive Stimuli and Pure Tones," *J. Acoust Soc. Am.*, Vol. 32, 1960, pp. 685–692.
46. W. A. Yost, F. L. Wightman, and D. M. Green, "Lateralization of Filtered Clicks," *J. Acoust Soc. Am.*, Vol. 50, 1971, pp. 1526–1530.
47. E. R. Hafter and R. H. Dye, Jr. "Detection of Interaural Differences of Time in Trains of High-Frequency Clicks as a Function of Interclick Interval and Number," *J. Acoust Soc. Am.*, Vol. 73, 1983, pp. 644–651.
48. E. R. Hafter and J. DeMaio, "Difference Thresholds for Interaural Delay," *J. Acoust Soc. Am.*, Vol. 57, 1975, pp. 181–187.
49. J. C. R. Licklider, J. C. Webster, and J. M. Hedlun, "On the Frequency Limits of Binaural Beats," *J. Acoust Soc. Am.*, Vol. 22, 1950, pp. 468–473.
50. D. R. Perrott and M. A. Nelson, "Limits for the Detection of Binaural Beats," *J. Acoust Soc. Am.*, Vol. 46, 1969, pp. 1477–1481.
51. S. Kuwada and T. C. T. Yin, "Binaural Interaction in Low-Frequency Neurons in Inferior Colliculus of Cat. I. Effects of Long Interaural Delays, Intensity, and Repetition Rate on Interaural Delay Function," *J. Neurophysiol.*, Vol. 50, 1983, pp. 981–999.
52. T. C. T. Yin and S. Kuwada, "Binaural Interaction in Low-Frequency Neurons in Inferior Colliculus of Cat. II. Effects of Changing Rate and Duration of Interaural Phase," *J. Neurophys.*, Vol. 50, 1983, pp. 1000–1019.
53. D. McFadden and E. G. Pasanen, "Binaural Beats at High Frequencies," *Science*, Vol. 190, 1975, pp. 394–396.
54. A. W. Mills, "Lateralization of High-Frequency Tones," *J. Acoust Soc. Am.*, Vol. 32, 1960, pp. 132–134.
55. D. W. Grantham and J. B. Ahlstrom, "Interaural Intensity Discrimination of Noise as a Function of Center Frequency, Duration, and Interaural Correlation," *J. Acoust. Soc. Am.*, Vol. 71, 1982, pp. 586–587(A).
56. E. R. Hafter, R. H. Dye, J. M. Nuetzel, and H. Aronow, "Difference Thresholds for Interaural Intensity," *J. Acoust Soc. Am.*, Vol. 61, 1977, pp. 829–834.
57. B. McA. Sayers, "Acoustic-Image Lateralization Judgments with Binaural Tones," *J. Acoust Soc. Am.*, Vol. 36, 1974, pp. 923–926.
58. W. A. Yost, "Lateral Position of Sinusoids Presented with Interaural Intensive and Temporal Differences," *J. Acoust Soc. Am.*, Vol. 70, 1981, pp. 397–409.
59. R. H. Whitworth and L. A. Jeffress, "Time versus Intensity in the Localization of Tones," *J. Acoust Soc. Am.*, Vol. 33, 1961, pp. 925–929.
60. R. H. Domnitz and H. S. Colburn, "Lateral Position and Interaural Discrimination," *J. Acoust Soc. Am.*, Vol. 61, 1977, pp. 1586–1598.
61. L. R. Bernstein and C. Trahiotis, "Lateralization of Low-Frequency, Complex Waveforms: The Use of Envelope-Based Temporal Disparities," *J. Acoust Soc. Am.*, Vol. 77, 1985, pp. 1868–1880.
62. J. L. Schiano, C. Trahiotis, and L. R. Bernstein, "Lateralization of Low-Frequency Tones and Narrow Bands of Noise," *J. Acoust Soc. Am.*, Vol. 79, 1986, pp. 1563–1570.
63. W. A. Yost and E. R. Hafter, "Lateralization," in W. A. Yost and G. Gourevitch (Eds.), *Directional Hearing*, Springer-Verlag, New York, 1987.
64. L. R. Bernstein and C. Trahiotis, "Lateralization of Sinusoidally Amplitude-Modulated Tones: Effects of Spectral Locus and Temporal Variation," *J. Acoust Soc. Am.*, Vol. 78, 1985, pp. 514–523.
65. E. R. Hafter and L. A. Jeffress, "Two-Image Lateralization of Tones and Clicks," *J. Acoust Soc. Am.*, Vol. 44, 1968, pp. 563–569.
66. G. Moushegian and L. A. Jeffress, "Role of Interaural Time and Intensity Differences in the Lateralization of Low-Frequency Tones," *J. Acoust Soc. Am.*, Vol. 31, 1959, pp. 1441–1445.
67. E. R. Hafter and S. C. Carrier, "Binaural Interaction in Low-Frequency Stimuli: The Inability to Trade Time and Intensity Completely," *J. Acoust Soc. Am.*, Vol. 51, 1972, pp. 1852–1862.
68. W. S. Woods and H. S. Colburn, "Test of a Model of Auditory Object Formation Using Intensity and ITS Discrimination," *J. Acoust Soc. Am.*, Vol. 91, 1992, pp. 2894–2902.
69. M. A. Stellmack and R. H. Dye, Jr. "The Effect of Onset Asynchronies between Components on the Ability to Detect a Delayed Component Embedded in a Diotic Complex," *J. Acoust Soc. Am.*, Vol. 86, 1989, p. S11.

70. C. Trahiotis and L. R. Bernstein, "Detectability of Interaural Delays over Select Spectral Regions: Effects of Flanking Noise," *J. Acoust Soc. Am.*, Vol. 87, 1990, pp. 810–813.

71. T. N. Buell and E. R. Hafter, "Combination of Binaural Information across Frequency Bands," *J. Acoust Soc. Am.*, Vol. 90, 1991, pp. 1894–1900.

72. L. R. Bernstein, "Spectral Interference in a Binaural Detection Task," *J. Acoust Soc. Am.*, Vol. 89, 1991, pp. 1306–1313.

73. R. H. Dye, Jr. "Combination of Interaural Information across Frequencies: Lateralization on the Basis of Interaural Delay," *J. Acoust Soc. Am.*, Vol. 88, 1990, pp. 2159–2170.

74. L. A. Jeffress, H. C. Blodgett, T. T. Sandel, and C. L. Wood, III, "Masking of Tonal Signals," *J. Acoust Soc. Am.*, Vol. 28, 1956, pp. 416–426.

75. L. A. Jeffress, "Binaural Signal Detection: Vector Theory," in J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory*, Vol. 2, Academic, New York, 1972.

76. D. E. Robinson and L. A. Jeffress, "Effects of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals," *J. Acoust Soc. Am.*, Vol. 35, 1963, pp. 1947–1952.

77. D. E. Robinson and J. P. Egan, "Lateralization of an Auditory Signal in Correlated Noise and in Uncorrelated Noise as a Function of Frequency," *Percept. Psychophys.*, Vol. 15, 1974, pp. 281–284.

78. T. L. Langford and L. A. Jeffress, "Effect of Noise Crosscorrelation on Binaural Signal Detection," *J. Acoust Soc. Am.*, Vol. 36, 1964, pp. 1455–1458.

79. I. J. Hirsh and M. Burgeat, "Binaural Effects in Remote Masking," *J. Acoust Soc. Am.*, Vol. 30, 1958, pp. 827–832.

80. W. T. Bourbon, Jr., "Effects of Bandwidth and Level of Masking Noise on Deduction of Homophasic and Antiphasic Tonal Signals," Ph.D. Dissertation, University of Texas, Austin, 1966.

81. P. J. Metz, G. von Bismark, and N. I. Durlach, "Further Results on Binaural Unmasking and the EC Model. II. Noise Bandwidth and Interaural Phase," *J. Acoust Soc. Am.*, Vol. 43, 1968, pp. 1085–1091.

82. H. C. Blodgett, L. A. Jeffress, and R. W. Taylor, "Relation of Masked Threshold to Signal-Duration for Various Interaural Phase-Combinations," *Am. J. Psychol.*, Vol. 71, 1958, pp. 283–290.

83. D. M. Green, "Interaural Phase Effects in the Masking of Signals of Different Durations," *J. Acoust Soc. Am.*, Vol. 39, 1966, pp. 720–724.

84. E. R. Hafter and S. C. Carrier, "Masking-Level Differences Obtained with a Pulsed Tonal Masker," *J. Acoust. Soc. Am.*, Vol. 47, 1970, pp. 1041–1047.

85. I. J. Hirsh, "The Relation between Localization and Intelligibility," *J. Acoust Soc. Am.*, Vol. 22, 1950, pp. 196–200.

86. H. Levitt and L. R. Rabiner, "Binaural Release from Masking for Speech and Gain in Intelligibility," *J. Acoust Soc. Am.*, Vol. 42, 1967, pp. 601–608.

87. M. B. Gardner, "Historical Background of the Haas and/or Precedence Effect," *J. Acoust Soc. Am.*, Vol. 43, 1968, pp. 1243–1248.

88. P. M. Zurek, "The Precedence Effect," in W. A. Yost and G. Gourevitch (Eds.), *Directional Hearing*, Springer-Verlag, New York, 1987.

89. H. Wallach, E. B. Newman, and M. R. Rosenzweig, "The Precedence Effect in Sound Localization," *Am. J. Psychol.*, Vol. 52, 1949, pp. 315–336.

90. J. V. Tobias and E. D. Schubert, "Effective Onset Duration of Auditory Stimuli," *J. Acoust Soc. Am.*, Vol. 31, 1959, pp. 1595–1605.

91. S. M. Abel and H. Kunov, "Lateralization Based on Interaural Phase Differences: Effects of Frequency, Amplitude, Duration, and Shape of Rise/Decay," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 955–960.

92. T. N. Buell, C. Trahiotis, and L. R. Bernstein, "Lateralization of Low-Frequency Tones: Relative Potency of Gating and Ongoing Interaural Delay," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 3077–3085.

93. B. Rakerd and W. M. Hartmann, "Localization of Sounds in Rooms II: The Effects of a Single Reflecting Surface," *J. Acoust Soc. Am.*, Vol. 78, 1985, pp. 524–533.

94. D. R. Perrott, K. Marlborough, P. Merrill, and T. Z. Strybel, "Minimum Audible Angle Thresholds Obtained under Conditions in Which the Precedence Effect is Assumed to Operate," *J. Acoust Soc. Am.*, Vol. 85, 1989, pp. 282–288.

95. E. R. Hafter, K. Saberi, E. R. Jensen, and F. Briolle, "Localization in an Echoic Environment," in Y. Cazals, L. Demany, and K. Horner (Eds.), *Auditory Physiology and Perception*, Pergamon, Oxford, 1992.

96. T. Houtgast and R. Plomp, "Lateralization Threshold of a Signal in Noise," *J. Acoust Soc. Am.*, Vol. 44, 1968, pp. 807–812.

97. J. V. Tobias and S. Zerlin, "Lateralization Threshold as a Function of Stimulus Duration," *J. Acoust Soc. Am.*, Vol. 31, 1959, pp. 1594–1595.

98. G. L. Ricard and E. R. Hafter, "Detection of Interaural Time Differences in Short-Duration Low-Frequency Tones," *J. Acoust Soc. Am.*, Vol. 53, 1973, pp. 335(A).

99. E. R. Hafter, R. H. Dye, Jr., and E. Wenzel, "Detection of Interaural Differences of Intensity in Trains of High-Frequency Clicks as a Function of Interclick Interval and Number," *J. Acoust Soc. Am.*, Vol. 73, 1983, pp. 644-651.

100. E. R. Hafter, T. N. Buell, and V. M. Richards, "Onset-Coding in Lateralization: Its Form, Site and Function," in G. M. Edleman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function: Neurobiological Bases of Hearing*, Wiley, New York, 1988.

101. E. R. Hafter and T. N. Buell, "Restarting the Adapted Binaural System," *J. Acoust Soc. Am.*, Vol. 88, 1990, pp. 806–812.

102. J. Blauert and J-P. Col, "Irregularities in the Precedence

Effect," in Y. Cazals, L. Demany, and K. Horner (Eds.), *Auditory Physiology and Perception*, Pergamon, Oxford, 1992.

103. R. K. Clifton, "Breakdown of Echo Suppression in the Precedence Effect," *J. Acoust Soc. Am.*, Vol. 82, 1987, pp. 1834–1835.
104. H. Lehnert and J. Blauert, "Virtual Auditory Environment," *Proc. 5th Int. Conf. Adv. Rob. IEEE-ICAR*, Vol. 1, 1991, pp. 211–216.
105. E. M. Wenzel, "Localization in Virtual Acoustic Displays," *Presence*, Vol. 1, 1992, pp. 80–107.
106. C. Pösselt, J. Schröeter, M. Optiz, P. L. Divenyi, and J. Blauert, "Generation of Binaural Signals for Research and Home Entertainment," *Proc. Int. Cong. Acoust.*, Vol. 1, B1–6, Toronto, 1987.
107. M. R. Schroeder, "Improved Quasi-Stereophony and 'Colorless' Artificial Reverberation," *J. Acoust Soc. Am.*, Vol. 33, 1961, pp. 1061–1064.
108. D. W. Grantham, "Detection and Discrimination of Simulated Motion of Auditory Targets in the Horizontal Plane," *J. Acoust Soc. Am.*, Vol. 79, 1986, pp. 1939–1949.
109. E. R. Hafter and F. Briolle, "Localization of Complex Sounds in the Presence of Echoes," *Proc. 13th Int. Cong. Acoust.*, 1989, pp. 495–498.
110. M. Morimoto and Y. Ando, "On the Simulation of Sound Localization," *J. Acoust Soc. Jpn. (E)*, Vol. 1, 1980, pp. 167–174.
111. F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening I: Stimulus Synthesis," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 858–867.
112. N. I. Durlach, R. Rigopulous, X. D. Pang, B. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, "On the Externalization of Auditory Images," *Presence*, Vol. 1, 1992, pp. 251–257.
113. J. M. Loomis, C. Hebert, and J. G. Cicinelli, "Active Localization of Virtual Sounds," *J. Acoust Soc. Am.*, Vol. 88, 1990, pp. 1757–1764.
114. J. A. Altman and O. V. Viskov, "Discrimination of Perceived Movement Velocity for Fused Auditory Image in Dichoic Stimulation," *J. Acoust Soc. Am.*, Vol. 61, 1977, pp. 816–819.
115. D. T. Perrott and A. B. Musicant, "Minimum Auditory Movement Angle: Binaural Localization of Moving Sound Sources," *J. Acoust Soc. Am.*, Vol. 62, 1977, pp. 1463–1466.
116. D. W. Grantham and F. L. Wightman, "Detectability of Varying Interaural Temporal Differences," *J. Acoust Soc. Am.*, Vol. 63, 1978, pp. 511–523.
117. K. Saberi and D. R. Perrott, "Minimum Audible Movement Angles as a Function of Sound Source Trajectory," *J. Acoust Soc. Am.*, Vol. 88, 1990, pp. 2639–2644.
118. D. W. Grantham, "Discrimination of Dynamic Interaural Intensity Differences," *J. Acoust Soc. Am.*, Vol. 76, 1984, pp. 71–76.
119. L. A. Jeffress, "A Place Mechanism of Sound Localization," *J. Comp. Physiol. Psychol.*, Vol. 41, 1948, pp. 35–39.
120. J. E. Rose, N. B. Gross, C. D. Geisler, and J. E. Hind, "Some Neural Mechanisms in the Inferior Colliculus of the Cat Which May be Relevant to Localization of a Sound Source," *J. Neurophys.*, Vol. 29, 1966, pp. 288–314.
121. J. N. Goldberg and P. B. Brown, "Functional Organization of the Dog Superior Olivary Complex: An Anatomical and Electrophysiological Study," *J. Neurophys.*, Vol. 31, 1968, pp. 639–656.
122. C. Carr and M. Konishi, "A Circuit for Detection of Interaural Time Differences in the Brainstem of the Barn Owl," *J. Neurosci.*, Vol. 10, 1990, pp. 3227–3246.
123. E. M. Overholt, E. W. Rubel, and R. L. Hyson, "A Circuit for Coding Interaural Time Differences in the Chick Brainstem," *J. Neurosci.*, Vol. 12, 1992, pp. 1698–1708.
124. H. S. Colburn, "Theory of Binaural Interaction Based on Auditory-Nerve Data II. Detection of Tones in Noise," *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 525–533.
125. R. M. Stern and H. S. Colburn, "Lateral-Position-Based Models of Interaural Discrimination," *J. Acoust Soc. Am.*, Vol. 77, 1985, pp. 753–755.
126. K. J. Gabriel and H. S. Colburn, "Interaural Correlation Discrimination: I. Bandwidth and Level Dependence," *J. Acoust. Soc. Am.*, Vol. 69, 1981, pp. 1394–1401.
127. E. Osman, "A Correlation Model of Binaural Masking Level Differences," *J. Acoust Soc. Am.*, Vol. 50, 1971, pp. 1494–1511.
128. J. Blauert and W. Cobben, "Some Considerations of Binaural Cross-Correlation Analysis," *Acustica*, Vol. 39, 1978, pp. 96–103.
129. R. M. Stern, A. S. Zeiberg, and C. Trahiotis, "Lateralization of Complex Binaural Stimuli: A Weighted Image Model," *J. Acoust Soc. Am.*, Vol. 84, 1988, pp. 156–165.
130. T. A. Takahashi and M. Konishi, "Selectivity for Interaural Time Differences in the Owl's Midbrain," *J. Neurosci.*, Vol. 6, 1986, pp. 3413–3422.
131. H. Wagner, T. T. Takahashi, and M. Konishi, "Representation of Interaural Time-Differences in the Central Nucleus of the Barn Owl's Inferior Colliculus," *J. Neurosci.*, Vol. 7, 1987, pp. 3105–3116.
132. R. M. Stern and C. Trahiotis, "The Role of Consistency of Interaural Timing over Frequency in Binaural Lateralization," in Y. Cazals, L. Demanyand, and K. Horner (Eds.), *Auditory Physiology and Perception*, Pergamon, Oxford, 1992.
133. W. Lindemann, "Extension of a Binaural Cross-Correlation Model by Contralateral Inhibition. II. The Law of the First Wavefront," *J. Acoust Soc. Am.*, Vol. 80, 1986, pp. 1623–1630.
134. W. Lindemann, "Extension of a Binaural Cross-Correlation Model by Contralateral Inhibition. I. Simulation of

Lateralization for Stationary Signals," *J. Acoust Soc. Am.*, Vol. 80, 1986, pp. 1608–1622.

135. W. A. van Bergeijk, "Variation on a Theme of von Bekesy: A Model of Binaural Interaction," *J. Acoust Soc. Am.*, Vol. 34, 1962, pp. 1431–1437.
136. R. M. Stern and H. S. Colburn, "Theory of Binaural Interaction Based on Auditory-Nerve Data IV. A Model for Subjective Lateral Position," *J. Acoust Soc. Am.*, Vol. 64, 1978, pp. 127–140.
137. F. A. Webster, "The Influence of Interaural Phase on Masked Thresholds I. The Role of Interaural Time Deviation," *J. Acoust Soc. Am.*, Vol. 23, 1951, pp. 452–462.
138. N. I. Durlach, "Equalization and Cancellation Theory of Binaural Masking-Level Differences," *J. Acoust Soc. Am.*, Vol. 35, 1963, pp. 452–462.
139. N. I. Durlach, "Binaural Signal Detection: Equalization and Cancellation Theory," in J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory*, Vol. II, Academic, New York, 1972.
140. R. H. Domnitz and H. S. Colburn, "Analysis of Binaural Detection Models for Dependence on Interaural Target Parameters," *J. Acoust Soc. Am.*, Vol. 59, 1977, pp. 591–601.
141. R. E. Wicksberg and D. Oertel, "Delayed, Frequency-Specific Inhibition in the Cochlear Nuclei of Mice: A Mechanism for Monaural Echo Suppression," *J. Neurosci.*, Vol. 10, 1990, pp. 1762–1768.
142. E. B. Young, C. A. Spiro, J. J. Rice, and H. F. Voigt, "Neural Organization and Responses to Complex Stimuli in the Dorsal Cochlear Nucleus," *Philos. Trans. Roy. Soc. Lond. B*, Vol. 336, 1992, p. 407.
143. M. R. Schroeder, "Improved Quasi-Stereophony and 'Colorless' Artificial Reverberation," *J. Acoust Soc. Am.*, Vol. 33, 1961, pp. 1061–1064.
144. S. Shamma, N. Shen, and P. Gopalaswany, "Binaural Processing without Neural Delays," in H. Duifhuis, J. W. Horst, and H. P. Wit (Eds.), *Basic Issues in Hearing*, Academic, London, 1988.
145. C. S. Watson and B. T. Mittler, "Time-Intensity Equivalence in Auditory Lateralization," *Psychon. Sci.*, Vol. 2, 1965, pp. 219–220.

118

LOUDNESS

BERTRAM SCHARF

1 INTRODUCTION

Loudness is the subjective magnitude or intensity of sound. (The word *intensity* is used here in its generic sense and is akin to the word *volume* as used by many engineers. Psychoacoustics reserves the word volume to designate the perceived size of a sound.) Every sound can be characterized as located on a scale of loudness going from nearly inaudible to intolerably loud. In English, *loudness* and its associated adjective *loud* are common, everyday words that apply directly to auditory experience. This readily available vocabulary has facilitated the psychoacoustical measurement of loudness. Other languages have coined special words for loudness, for example, *sonie* in French, *Lautheit* in German.

Historically, the measurement of loudness has involved three distinct tasks for listeners. One task has been to scale the loudness of a series of sounds, usually as a function of sound pressure level (SPL); another task has been to match for loudness sounds that differ along a single dimension, such as spectral content or duration; a third task is to discriminate between two sounds that differ only with respect to level. This chapter distills a large number of experiments devoted to one or another of these three types of measurement. The first two tasks are treated together, because *loudness functions* obtained by scaling different sounds (such as a tone and white noise) can be related to one another by a matching operation and matching, in turn, gives rise to *equal-loudness contours*. The third task is treated separately; it provides measures of *loudness discrimination*, the sensitivity of the auditory system to changes or differences in level.

Loudness functions and contours are related to stimulus variables (sound pressure level, frequency, bandwidth, duration, temporal properties, background) and to listener variables (listening mode, whether monaural or binaural, prior sound exposure, auditory pathology). Discrimination is also related to many of these same variables. After the presentation of the basic data, psychophysical and physiological models of loudness and of loudness discrimination are discussed briefly.

2 LOUDNESS FUNCTIONS AND EQUAL-LOUDNESS CONTOURS

This section begins with the loudness function for a 1-kHz tone. This function, which relates loudness to sound pressure, is the standard used for the measurement of the loudness of all other sounds. Later sections show how the loudness of pure tones at other frequencies depends on SPL, how the loudness of complex sounds varies with bandwidth, how loudness depends on duration and other temporal properties, and how it is affected by the presence of other sounds.

2.1 Sound Pressure (Loudness Function for a 1-kHz Tone)

Figure 1 gives the loudness of a 1-kHz tone as a function of SPL. The straight line, including the dashed section, is from the international standard[1] and has a slope of 0.6, which means that for every 10-dB increase in level, loudness doubles. The straight line on these double-logarithmic coordinates describes a power function

$$L = kP^{0.6}, \qquad (1)$$

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

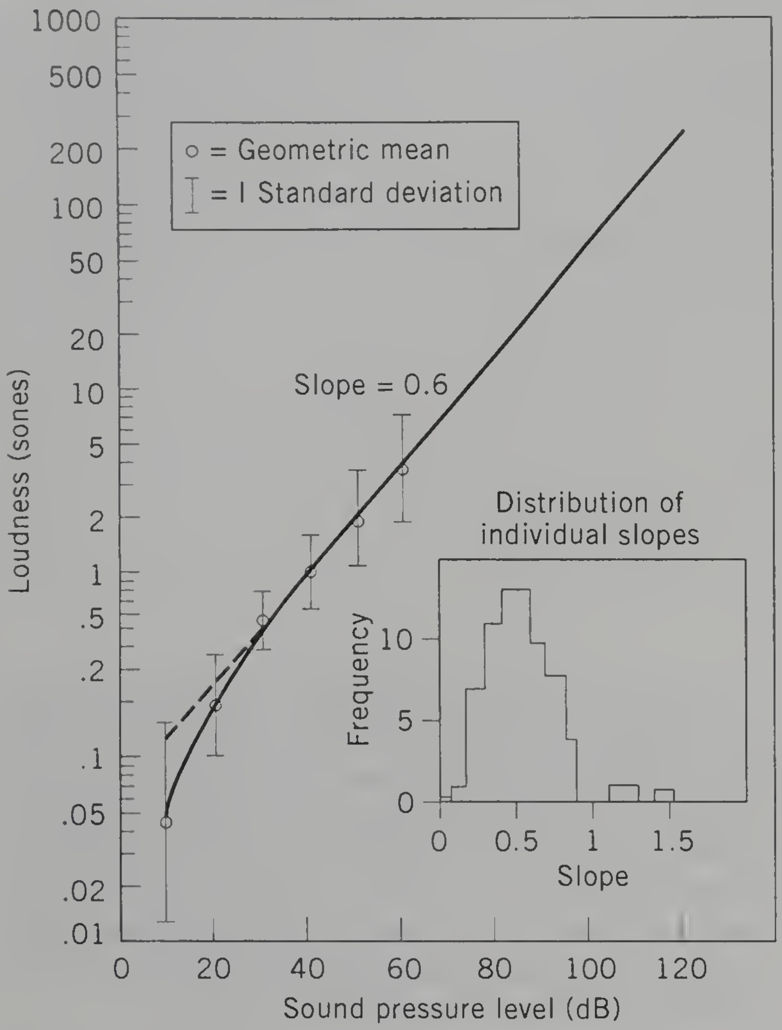


Fig. 1 Loudness function for a binaural, 1000-Hz tone. Loudness in sones is plotted as a function of the SPL in decibels measured at the center of the locus of the listener's head with the listener absent. The straight portion of the solid line is from the international standard for a 1-kHz tone[1] and has a slope of 0.6. The curved portion was taken from Ref. 26 and is based on a number of studies near threshold. The loudness of a 1000-Hz tone increases with an exponent of 0.6 according to the power law $L = 0.01(P - P_0)^{0.6}$, where L is loudness in sones, P is sound pressure in micropascals, and P_0 has the value of 45 μPa, which is the effective threshold at which loudness begins to increase with sound pressure; the subtractive constant is included to reflect the curvature of the loudness function at low levels. The slope of the straight line drawn in log–log coordinates is the exponent of the power function. An exponent of 0.6 means that loudness doubles whenever SPL increases by 10 dB. The data in the figure are the geometric means of magnitude estimates by 70 naive listeners who were run in groups of 4–11 in an anechoic room with the loudspeaker at a distance of about 3 m[7]. Curves fitted to individual data by the method of least squares yielded the 70 slopes plotted in the inset. Their mean is 0.6.

where L is loudness in sones, $k = 0.01$, and P is sound pressure in micropascals. The unit for loudness, the *sone*, is defined such that the loudness of a binaural 1-kHz tone at 40 dB SPL is equal to 1 sone. The standard loudness function was based on many investigations, most requiring listeners to evaluate loudness directly, by assigning a number or a ratio to the loudness of a tone. However, none of those investigations had made detailed measurements near threshold. Careful examination[2] of the function below about 40 dB sensation level (SL: the level above an individual listener's threshold or the SPL of the presented sound minus the SPL for that sound at the listener's threshold) showed that loudness changes more rapidly with level below about 30 dB SPL than above, as indicated by the solid line that curves down below 30 dB in Fig. 1. One way to modify the power law so as to represent this curvature is to deduct the effective threshold P_0 from P in Eq. (1). The modified power function is then

$$L = k(P - P_0)^{0.6}. \tag{2}$$

The effective threshold is that level at which the sound begins to have a sensory effect in the experimental context. A value for P_0 of 45 μPa is a reasonable value, the equivalent of a few decibels above the expected detection threshold for a binaural 1-kHz tone. However, other modifications to Eq. (1) can also yield a reasonable function.[3] One of these modifications, proposed by Zwislocki[4] (also see a closely related proposal by Humes and Jesteadt[5]), has the advantage that it fits loudness data when thresholds have been elevated by hearing impairment[6] or masking. Whatever the best fit, the important point is that the loudness function is steeper below approximately 30 dB than above.

Figure 1 also presents a set of data from 70 listeners who judged a 1-kHz tone presented from a loudspeaker in an anechoic room.[7] The circles are the geometric means for the 70 naive listeners who were run in groups of 4 to 11. The inset shows the distribution of the 70 slopes based on each listener's individual curve that had been determined by the method of least squares. The mean of these individual slopes is 0.6. The large variability in the individual slopes almost surely reflects largely judgmental differences among the listeners rather than sensory differences. If the individual slopes represented the true course of loudness, then the listener with the steepest slope of 1.5 would hear people's voices at loudnesses many thousand times greater than would the listener with the shallowest slope of 0.1. (To arrive at this relation, we need only assume that the two listeners have similar thresholds and similar loudness sensations near threshold.)

The function in Fig. 1 is defined for a tone presented from a loudspeaker in a free field, and the data shown were obtained under those listening conditions. The same function (except for a small threshold correction) applies to a tone presented binaurally through a pair of earphones. To apply to monaural presentation through a sin-

TABLE 1 Sone Values for a Binaural 1-kHz Tone and for a Binaural White Noise Listed as a Function of Sound Pressure Level

SPL (dB)	Tone	Noise	SPL (dB)	Tone	Noise	SPL (dB)	Tone	Noise
10	0.052	—	50	2.00	3.85	90	32.0	46.0
12	0.072	—	52	2.30	4.45	92	36.8	50.5
14	0.095	—	54	2.64	5.20	94	42.2	57.5
15	0.110	—	55	2.83	5.60	95	45.3	61.0
16	0.125	—	56	3.03	6.00	96	48.5	65.0
18	0.155	—	58	3.48	7.00	98	55.7	72.0
20	0.190	—	60	4.00	7.85	100	64.0	80
22	0.230	—	62	4.59	8.9	102	73.5	91
24	0.280	—	64	5.28	10.2	104	84.4	102
25	0.305	—	65	5.66	10.9	105	90.5	108
26	0.330	—	66	6.06	11.5	106	97.0	114
28	0.395	.450	68	6.96	13.0	108	111	128
30	0.460	.580	70	8.00	14.7	110	128	
32	0.550	.720	72	9.19	16.4	112	147	
34	0.640	.900	74	10.6	18.5	114	169	
35	0.700	1.00	75	11.3	19.5	115	181	
36	0.750	1.10	76	12.1	20.6	116	194	
38	0.860	1.36	78	13.9	23.2	118	223	
40	1.00	1.65	80	16.0	26.0	120	256	
42	1.15	2.00	82	18.4	29.0			
44	1.32	2.40	84	21.1	32.5			
45	1.41	2.60	85	22.6	34.8			
46	1.52	2.80	86	24.3	36.5			
48	1.74	3.28	88	27.9	41.0			

Note: A value of 1 sone is assigned to the loudness of a binaural 1-kHz tone at 40 dB SPL. Above 1 sone, the values listed for the tone are from the international standard (Ref. 1). Below 1 sone, the values are based on the curve in Fig. 1, which summarizes a number of studies of loudness at low intensities. The sone values for a white noise are based on measurements of the loudness of white noise. (Adapted with permission from Ref. 26.)

gle phone, the function would have to be moved to the right to take into account the higher monaural threshold and the reduced monaural loudness, as described below.

Table 1 presents the values of the solid-line function of Fig. 1. For a 1000-Hz tone, the sone value is given at even values of SPL from 10 to 120 dB and also at all levels ending with the digit 5. (Table 1 also gives the sone values for white noise.)

2.2 Frequency

The loudness function at 1 kHz is not valid at all frequencies. As can be seen from the curve relating threshold in the quiet to frequency (Chapter 123), our ears are most sensitive to the middle frequencies, including 1 kHz. When the level is increased above threshold, frequencies do not suddenly become equally effective and produce the same loudness. However, the differences among the lower frequencies do become smaller as level increases. These differences are quantified by having listeners match the loudness of tones at other frequencies to that of a tone at 1 kHz or by having them estimate loudness at other frequencies as a function of level, just as has been done for a 1-kHz tone. Much effort has been directed toward measuring the equal-loudness contours both under earphones and in the free field, so these are presented first. Relatively few data are available for directly measured loudness functions at frequencies other than 1 kHz.

Although measured over half a century ago, the contours published by Fletcher and Munson[8] are still the standard for earphone listening. Figure 2 plots SPL as a function of frequency with loudness level as the parameter. *Loudness level* is the SPL at which a 1-kHz tone is judged as loud as another sound. The unit of loudness level is a decibel, but to distinguish it from ordinary decibels of SPL, it is called the *phon*. Fletcher and Munson's subjects matched in loudness 1-s tones at each of many other frequencies to a tone at 1 kHz. Thus, each contour represents a given loudness level, ranging from 13 to 113 phons. The bottom-most contour is the threshold curve that was at 3 phons for these subjects, whose average threshold at 1 kHz was 3 dB SPL. Other than on the threshold curve, all points on a contour represent equally

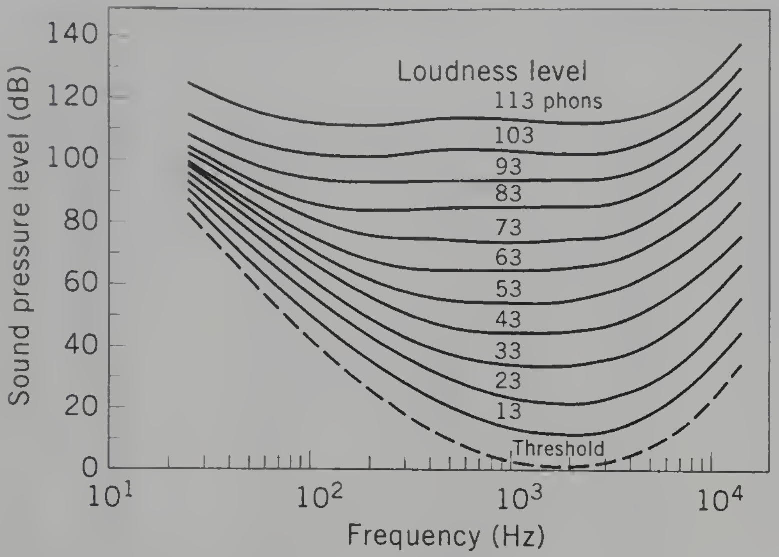


Fig. 2 Equal-loudness contours for pure tones presented through a pair of earphones. The ordinate gives the SPL at which a tone at the frequency given on the abscissa sounded equal in loudness to a 1-kHz reference tone. The level of the 1-kHz tone is the loudness level shown as the parameter on the curves. A method of constant stimuli was used in which the variable frequency was presented for 1 s followed 0.5 s later by the 1-kHz tone for 1 s. Rise–fall times were 100 ms. After two such presentations the subject responded whether the reference tone was louder or softer than the variable-frequency tone. Eleven experienced subjects each made 27 judgments at each loudness level for a total of 297 judgments, the median of which served in the construction of the curves in the figure. These curves were drawn through the median points and smoothed. The standard error of the combined results was between 1 and 2 dB, with the variability somewhat larger when the variable frequency was far from 1 kHz than when it was close. A given contour gives the combinations of frequency and SPL that yield the loudness level indicated on that contour. The bottom curve is based on threshold measurements obtained by a method of limits. Threshold at 1 kHz is assumed to be 3 dB. (Adapted from Ref. 18.)

loud tones. Thus on the 13-phon contour, a 100-Hz tone at 40 dB, a 1-kHz tone at 13 dB, and a 10-kHz tone at 33 dB are all equally loud. Most notable is the compression of the contours at the low frequencies, meaning that loudness grows most rapidly with sound pressure there; at higher frequencies the contours remain essentially parallel and loudness grows less rapidly. It is also noteworthy that the shape of the 13-phon contour is very similar to that of the threshold curve, although the two curves were obtained by very different procedures, one involving the detection of a tone, the other the equating in loudness of two tones alternating in time.

Equal-loudness contours have also been measured in the free field, with the tones presented from a loudspeaker. Figure 3 presents the contours measured by Robinson and Dadson[9] and adopted as an international standard.[10] The ordinate gives the SPL measured in the free field with the listener absent. The presence of the listener alters the sound field so that the sound pressure in the ear canal is greater around 4 kHz and smaller around 8 kHz than in earphone listening. Accordingly, the dip around 4 kHz and the peak near 8 kHz reflect primarily diffraction by the head and pinnae. Presumably, were the sound pressure at the eardrum measured in both earphone and free-field listening, the equal-loudness contours in Figs. 2 and 3 would be the same (except for an effect of the ear cushion, which may give rise to noise that reduces loudness of weak, low frequencies). (Some recent investigations have suggested that the contours in Fig. 3 should be higher at frequencies below 1 kHz; this would mean that to match the loudness of a 1-kHz tone, tones at the lower frequencies would have to be at higher levels than indicated in the figure.[11])

Having found a nice correspondence between threshold measurements and loudness matching, we now examine the relation between the outcome of loudness matching and direct estimations of loudness, which serve as the basis for the standard loudness function at 1 kHz.

From the equal-loudness contours for earphone listening (Fig. 2), we can calculate the loudness functions at frequencies other than 1 kHz. For earphone listening, we should expect loudness functions for frequencies above 1 kHz to be parallel to the 1-kHz function. Owing to threshold differences, the functions would be displaced to the left (lower thresholds) for frequencies between 1 and 4 kHz and to the right (higher thresholds) for frequencies above 4 kHz. On the other hand, functions for frequencies below 1 kHz will not be parallel because the contours are not parallel but are bunched together. For any given frequency, the loudness function can be calculated by taking the SPLs at which it is equal in loudness to a 1-kHz tone and the corresponding loudness in sones from Fig. 1 or Table 1. Some representative functions are shown in Fig. 4. These derived functions are in nice agreement with those measured by magnitude estimation at 100, 250, and 1000 Hz.[12] Schneider et al.[13] also reported reasonable agreement between directly measured loudness functions and published equal-loudness contours.

2.3 Bandwidth

The effect of signal bandwidth on loudness is revealed at its extreme by the difference between the loudness functions for a 1-kHz tone and for white noise. These differences can be gleaned from Table 1, which gives sone values for a 1-kHz tone and for white noise. At every level beginning at 28 dB SPL, a white noise is louder than an equally intense 1-kHz tone. At the midlevels, between about 50 and 60 dB, the noise is nearly twice as loud as a tone at the same level. In other terms, where the

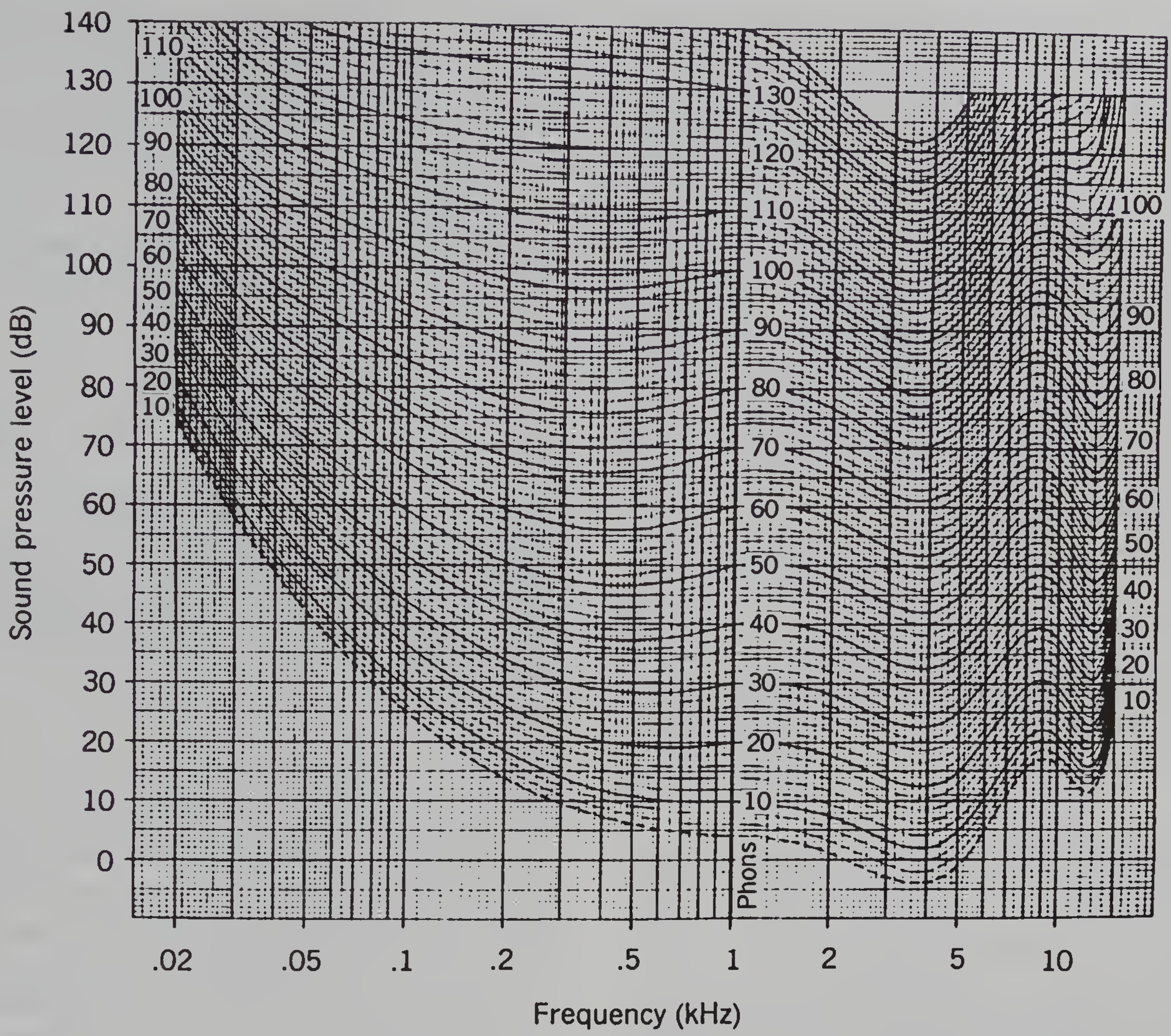


Fig. 3 Equal-loudness contours for pure tones presented in a free field. The ordinate, abscissa, and parameter are the same as in Fig. 2. Data were collected by the method of constant stimuli in a fashion similar to that described for Fig. 2. The results from a group of about 30 men and women aged 18–25 years were used to construct these contours, which constitute the international standard. Variability for the entire original group depended upon the difference in frequency between the variable tone and the reference 1-kHz tone and also on level. For moderate-frequency separations the standard deviation was about 5 phons, increasing to 10 phons for wide separations and also at high levels. The free-field contours differ from the earphone contours primarily at the high frequencies where the head, torso, and pinnae render the sound pressure generated at the ear canal entrance different from that measured in the sound field with the listener absent. Differences at low frequency are related more to leakage under the ear cushion and to greater influence of internal noise. (From Ref. 18, based on Ref. 9.)

difference is maximum, the tone must be 10 dB more intense than the noise to equal it in loudness; a 55-dB noise is just as loud as a 65-dB tone. Given that the difference between tone and noise depends on level, the white noise cannot follow the same simple power function above 30 dB that the tone does. Were the loudness function for white noise plotted on the double-logarithmic coordinates of Fig. 1, it would be bowed, with the loudness of noise growing more rapidly than that of the 1-kHz tone up to about 50 dB and more slowly above about 70 dB.

White noise contains all the audible frequencies and so is, for the human ear, the widest possible band of audible noise. When the bandwidth of the noise is narrowed, loudness begins to diminish even though the overall level of the noise is held constant. (However, given the nonlinear relationship between the loudness of white noise and the loudness of a 1-kHz tone, the change in loudness with bandwidth depends on overall level.) Figure 5 gives curves based on a matching procedure in which a 1-kHz tone was equated in loudness to a band of noise whose width is given on the abscissa. The level at which the tone was judged equal in loudness to the tone is the loudness level of the noise, shown on the

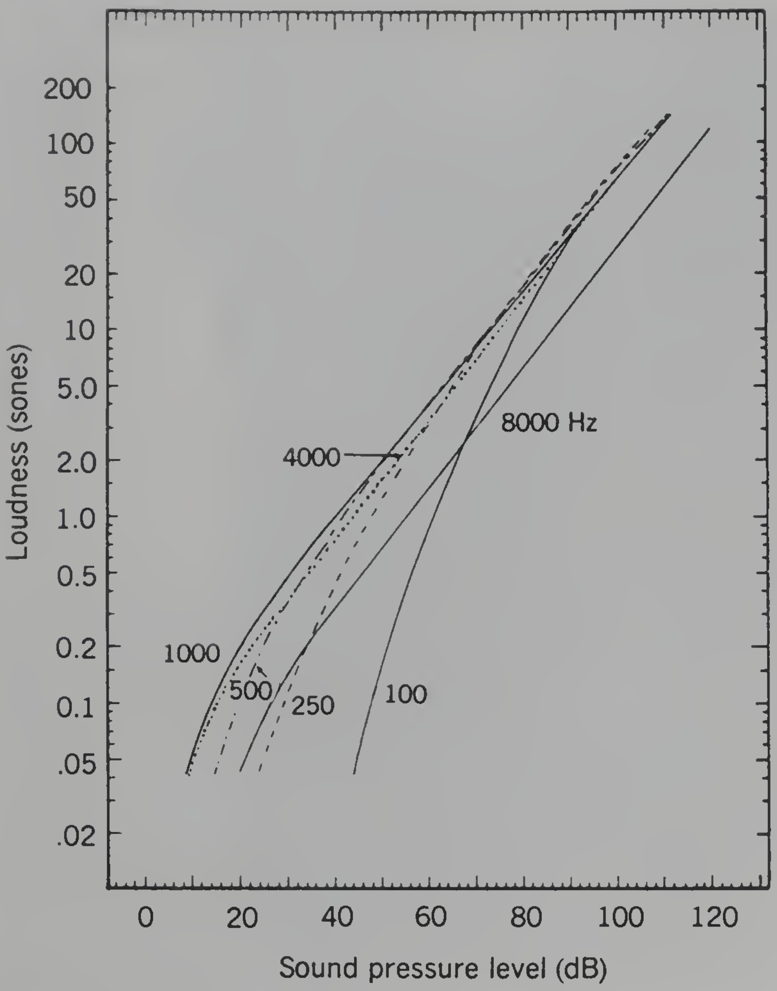


Fig. 4 Loudness of tones at different frequencies as a function of SPL. Functions are derived from the equal-loudness contours of Fig. 2 and the 1-kHz loudness functions of Fig. 1. (From Ref. 26.)

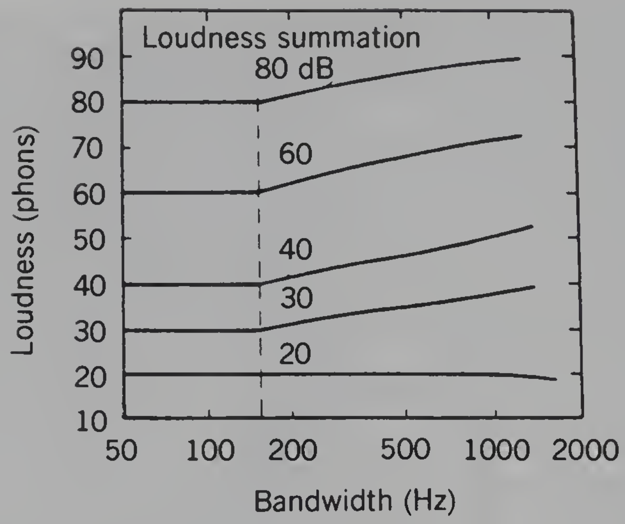


Fig. 5 Loudness level of a band of filtered white noise as a function of its width. Parameter on the curves is the overall SPL of the noise bands. The dashed line indicates the location of the critical band for these noises centered on 1000 Hz. (From Ref. 18.)

ordinate. Hence, the curves in Fig. 5 are not equal-loudness contours; rather, they show directly how loudness level changes as a function of bandwidth. The increase in loudness with bandwidth is called *loudness summation*. Loudness summation is greatest where the difference between the loudnesses of white noise and of a 1-kHz tone is greatest, around 60 dB SPL. At higher and lower levels, loudness increases less rapidly with bandwidth. Indeed, at 20 dB, loudness diminishes somewhat at very large bandwidths. The mechanisms underlying this complex set of relations are discussed in the section on models of loudness.

The increase in loudness with bandwidth does not begin until the *critical band* is exceeded. The critical band, which plays an essential role in masking (Chapter 115) and frequency analysis (Chapter 116), depends on the center frequency of the sound; it is 160 Hz at 1 kHz. At subcritical bandwidths, loudness does not change with width. However, at very narrow bandwidths, narrower than 40 or 50 Hz, fluctuations become so strong as to make loudness judgments difficult and highly variable. Indeed, at high frequencies and moderate levels, rapid amplitude fluctuations seem to enhance the loudness of very narrow bands of noise.[14]

The rules of loudness summation that apply to noises with continuous, flat spectra apply also to multitone complexes, that is, line spectra, sounds composed of two or more pure tones. Loudness depends primarily on the difference between the frequencies of the lowest and highest components, which expresses, in effect, the bandwidth of the multitone complex. Loudness appears to be independent of the number of components,[15,16] so that a two-tone or four-tone complex is as loud as an equally intense continuous-spectrum noise with the same bandwidth and overall level. Given a multitone complex, loudness is greatest when the components are evenly spaced across critical bands[16] and when they are set equally loud.[17]

In general, the loudness of a complex sound increases with bandwidth once the critical band has been exceeded. The increase in loudness is greatest at moderate SPLs. This level dependence means that the loudness of a broadband sound cannot be a power function of sound pressure, as shown above.

2.4 Duration

For the most part, loudness is independent of sound duration. However, the exceptions are important. At durations shorter than a few hundred milliseconds, loudness increases with sound duration. This increase is called *temporal summation* of loudness. At durations longer than a few seconds, loudness diminishes with duration up to 2 or 3 min but *only for weak* tones within about 30 dB of threshold or for tones at very high frequencies. This decrease is called *loudness adaptation.*

Temporal Summation A large number of studies have been devoted to measuring the dependence of loudness on durations ranging from a few milliseconds to a second or more. Results have been inconsistent, probably owing to the difficulty of judging loudness independently of perceived duration. Scharf and Houtsma[18] recently reviewed those studies. They suggested that the best summary of the data is the exponential

$$I(t) = \frac{I_\infty}{1 - e^{t/80}}, \qquad (3)$$

where $I(t)$ is the sound intensity required to maintain constant loudness, I_∞ is the asymptotic intensity at long durations, and t is stimulus duration in milliseconds. The time constant of 80 ms is based on an extensive round-robin study[19] completed in the 1970s in which a number of different laboratories repeated the same measurements; the value of 80 ms is reasonably close to most of the time constants that can be estimated from other, less extensive investigations. That study focused on the loudness of 1-kHz tone bursts and found that up to 100 ms loudness depends on total energy. Although the round-robin studies used only 1-kHz signals, other investigations do not reveal any consistent divergences at other frequencies or for noise.

According to Eq. (3), loudness is kept constant by reducing intensity in nearly direct proportion as duration lengthens up to near 80 ms. This trading relation suggests that at very brief durations loudness depends on sound energy and not on intensity (or sound pressure). It also means that loudness level comes to within 1 phon of its asymptotic value when duration reaches about 125 ms and to within half a phon near 180 ms.

Loudness Adaptation Many earlier reports in the literature purported to show that loudness adaptation is a general and strong phenomenon. Almost all of them were based on interaural loudness matching, which gave rise to interaural interaction and to induced adaptation. Studies that avoid such interactions show that after approximately 200 ms loudness generally does not change with sound duration. Only when a steady sound is within about 30 dB of threshold does loudness adapt. The closer to threshold, the greater the decline in loudness. At 5 dB SL, a tone becomes inaudible for many listeners. Tones above about 10 kHz show loudness adaptation already at much higher levels.[20] Noises and intermittent tones show considerably less loudness adaptation than do steady tones.[21]

Although most sounds above 30 dB SL do not decline in loudness over time, they can be induced to do so by incrementing them intermittently (e.g., every 2 s for 1 s or every 20 s for 10 s) or by presenting the steady sound monaurally and introducing an intermittent sound in the contralateral ear.[22] (See also Chapter 44 of Volume 2 on adaptation in the auditory system.)

2.5 Temporal Properties Other Than Duration

Rise–Fall Time A rapid onset and offset of a sound produces transients across a wide range of frequencies and thus enhances the loudness of the sound at most listening levels, especially if the sound is a brief, narrow-band stimulus such as a tone. The greater loudness results from the spread of energy beyond a single critical band. Once the rise–fall time exceeds a few milliseconds, it has a negligible effect on loudness.[23,24]

Temporal Pattern Kuwano et al.[24] reported small effects of envelope pattern on the loudness of impulsive sounds. Changes in the modulation rate or modulation frequency of a 1-kHz sinusoid resulted in loudness changes equivalent to less than 1 dB.

Intermittency Repeating a sound increases its loudness, provided, according to earlier studies from Germany, that the repetition rate is not less than 2 Hz.[25] At slower rates individual pulses are too far apart to permit temporal summation of loudness. Such an interpretation is in line with data on the loudness of double pulses for which time constants are of the order of 200 ms (see Ref. 26, pp. 210–211, for a brief review). However, a round-robin study conducted in 11 laboratories in Japan with 111 subjects suggests that loudness increases with the number of repetitions—all presented over nearly the same duration—whether the temporal separation between individual pulses is nearly 1 s or only 30 ms.[27] In other words, temporal summation seemed to be as good with a spacing of 1 s as with a spacing of 30 ms. The discrepancy with the earlier data may reflect cultural differences in the criteria used in judging loudness.

A somewhat different discrepancy relates to the temporal integration of the loudness of pulse trains that has been shown to continue to increase over durations up to and beyond 2 s.[28] Such integration times are much longer than those measured for continuous sounds. Berglund and Nordin[28] suggested that the off times in an intermittent sound may permit the longer integration times.

Time-Varying Sounds The assessment of sounds that vary in level and spectrum over long time periods requires continuous judgments. Namba[29] presented a model for such conditions and examples of various outcomes. It appears that the instantaneous loudness thus measured depends largely on the same stimulus and observer characteristics as a short-duration sound.[30] However, decreasing the level of a tone slowly over time provokes a more rapid decline in loudness than the physical decrease alone can account for, a kind of *decruitment*.[31,32] In this case, as in adaptation, the loudness depends not only on the immediate stimulus characteristics but also on the preceding history of the stimulus.

2.6 Background Sounds (Partial Masking)

When listeners are exposed to more than one sound at a time, they can judge the loudness of one sound to the exclusion of all other sounds or they can judge the overall loudness of all the sounds together. However, natural sounds coming from different directions with different time signatures as well as spectra are not readily judged together. Laboratory studies have concentrated on sounds presented via earphones with onset and offset disparities that permit the listener to judge one of two sounds despite strong interaction in the periphery of the auditory system. A typical set of results is shown in Fig. 6 for a 1-kHz tone presented and adjusted in level to match in loudness the same tone presented against a white noise.[33] The level of the tone in noise is given on the abscissa, and the level of the noise is the parameter on the curves. The noise was turned on for 2.8 s and the partially masked tone came on 0.4 s later and went off 1.9 s later, 0.5 s before the noise went off. During a 2.8-s interval with no noise, the tone was presented for 1.8 s and its level adjusted by the listener. The dashed line traces the levels at which both tones would be at the same level. All data points fall below the dashed line because for these levels of tone and noise the loudness of the tone was always reduced by the noise. This reduction in loudness caused by another sound is called *partial masking*. Hence, the level of the tone in noise is always greater than that of the equally loud tone in the quiet. However, the partial masking diminishes as the level of the tone in noise increases; the loudness level of the masked tone increases more rapidly than the dashed line, and the more intense the masker the more rapidly the loudness recovers. But even at the highest levels of the masked tone, some residual masking remains.

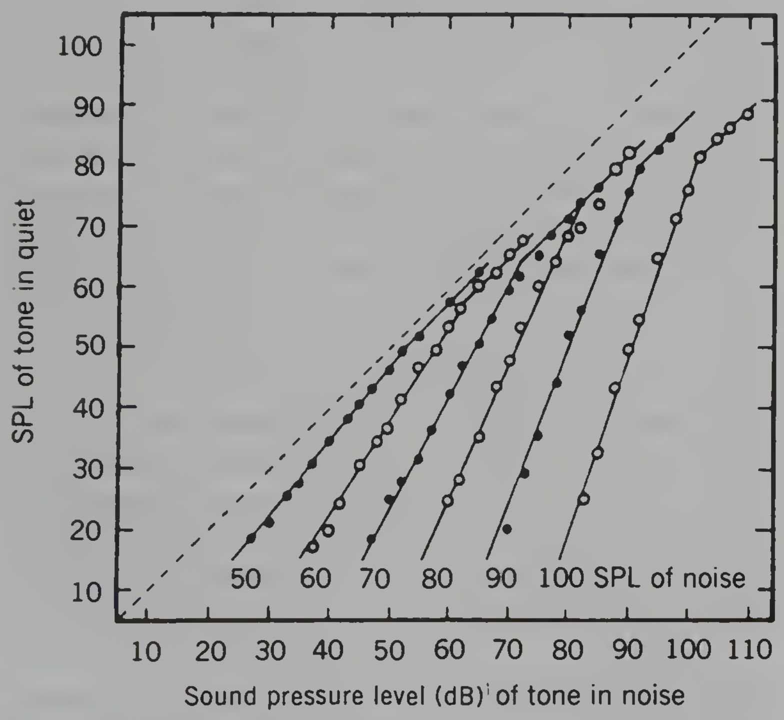


Fig. 6 Sound pressure level at which a 1-kHz tone in quiet is judged equal in loudness to a tone partially masked by white noise. Parameter on the curves is the level of the masking noise in decibels. Symbols are filled and unfilled for clarity. (From Ref. 26, based on Ref. 33.)

The rapid growth of loudness seen in Fig. 6 is called *loudness recruitment*; it is a characteristic of partial masking and also of sensorineural hearing impairment (see below). Loudness grows even more rapidly in the presence of narrow-band noise centered on the tone.[34,35] Whereas in wide-band noise, loudness grows rapidly up to 20–30 dB above masked threshold, in narrow-band noise it grows rapidly up to 10 or 15 dB above threshold, by which level the loudness almost always reaches its normal, unmasked loudness.

Partial masking, like complete masking, is greatest when the masker occupies the same frequency region as the signal.[36] Masking decreases when the masker and signal are in different frequency regions. For signals near masked threshold, loudness is reduced more by a masker at a lower frequency than the signal, just as in complete masking. For more intense signals, loudness is reduced more by a masker at a higher frequency.

2.7 Listener Variables

First data on binaural versus monaural loudness will be presented. Then, the effect on loudness of prior sound exposure and auditory pathology will be outlined. The available data on such factors as attention and individual differences do not warrant a treatment in this compendium.

Binaural versus Monaural Loudness Binaural and monaural loudness vary in the same way as a function of stimulus variables such as SPL, frequency, and bandwidth. The major difference is that a sound to both ears is judged to be nearly twice as loud as the same sound to one ear. Many studies based on equal-loudness matching and direct scaling support this conclusion. Nonetheless, the precise loudness ratio has not been agreed upon.[37] Agreement is better on the binaural advantage at threshold, which is 3 dB (Chapter 123). Equation (2) above for the binaural loudness function can be modified as follows to serve for the monaural function:

$$L = (0.01B^{-1}(P - 1.4P_0)^{0.6} \qquad (4)$$

where, as before, L is in sones, P is in micropascals, P_0 is the effective threshold with a value 45 μPa, and B is the binaural-to-monaural loudness ratio, which may have a value between 1.7 and 2.0.

In terms of decibels, above about 30 dB SL where the loudness function becomes straight, a monaural sound must be approximately 10 dB more intense than a binaural sound to be judged equally loud. Below 30 dB, as the loudness function steepens, the difference in decibels decreases to 3 dB at threshold, but the binaural-to-monaural loudness ratio remains fairly constant.

Prior Sound Exposure (Auditory Fatigue) Exposure to an intense sound for a short period or to a weaker sound for a longer period may result in a temporary reduction in the loudness of sounds presented after cessation of the fatiguing sound just as such exposure results in higher thresholds, the temporary threshold shift (TTS). The loudness of a sound a few decibels above the elevated threshold is less than that of the same sound prior to the exposure to the fatiguing sound. As the level of the post exposure test sound is increased, loudness grows rapidly, showing loudness recruitment like that noted in the presence of a partially masking sound. For example, as TTS increases from 10 to 30 dB, the slope of the post exposure loudness function goes from 0.62 to 1.06.[38] After exposure to an intense sound, TTS is greatest about half an octave above the frequency of the fatiguing sound, whereas temporary loudness shift (TLS) is greatest at the fatiguing frequency. After exposure to a moderately intense sound of 65 dB SPL, both TTS and TLS are greatest at the fatiguing frequency, and TLS is greater (measured in decibels) than TTS.[39]

Auditory Pathology Patients with cochlear pathology such as noise-induced hearing loss or Ménière's disease usually show loudness recruitment, a rapid increase in loudness as a function of SPL. This pathological recruitment looks much like that measured against noise in normally hearing listeners. Such patients with a hearing loss in only one ear would yield functions like those shown in Fig. 6, but instead of matching a tone in noise to a tone in quiet, the functions would be based on matching a tone in the healthy ear to a tone in the impaired ear. The parameter on the curves would be hearing loss instead of noise level, and the data would be from a group of impaired listeners. Just as the loudness function in noise steepens as the noise is made more intense (and hence the masked threshold higher), so does it steepen as the hearing loss becomes greater. Generally, loudness grows so rapidly above the elevated threshold that loudness in the impaired ear is about equal to that in a normal ear at about 30 dB SL and above.[6]

Loudness summation over bandwidth is also affected by cochlear pathology so that loudness increases less than in normal hearing when stimulus bandwidth widens.[40,41]

3 LOUDNESS DISCRIMINATION

Loudness discrimination (also referred to as level or intensity discrimination) is the ability to detect a difference in the loudness of two sounds. The many published reports on level discrimination permit us to specify the physical variables that control discrimination. However, subjects vary greatly in their ability to discriminate level, and the causes of this variability are unclear. This section summarizes the data on the role of level, frequency, spectral bandwidth, duration, and sound context in discrimination. First, the major procedural issues are addressed.

Two general stimulus paradigms have been employed in experiments on discrimination. In the typical *burst-comparison* paradigm, a sound is presented successively first at one level and then at another, and the listener's task is essentially to report which sound is louder. In the *modulation* paradigm, a single sound is increased briefly in level; the listener reports whether the increment was heard. Any of the several psychophysical procedures described in Chapter 115 can be applied under these stimulus paradigms. The comparison paradigm is a better measure of discrimination because it forces the listener to base his or her decision almost entirely on the difference, whereas modulation always includes a single onset and offset that may provide the basis for the listener's decision. (Both bursts in the comparison paradigm have essentially the same onsets and offsets.) In many ways, modulation is more akin to a masking paradigm since it requires the detection of one sound—here, the increment—in the presence of another—here, the base or pedestal. However, the modulation paradigm or increment detection is generally included in discussions of discrimination rather than of masking, and so a brief section is devoted to it after discussion of the comparison paradigm.

3.1 Burst Comparison

Level and Frequency Level discrimination for pure tones improves with sensation level. At 1 kHz, the level difference ΔL required to discriminate between two levels decreases from approximately 2 dB near threshold to approximately 0.5 dB at 80 dB SL.[42] This relationship holds at frequencies up to about 6 kHz; it can be summarized as follows:

$$\Delta L = 1.644\text{dB} - 0.0141 \times \text{SL}, \tag{5}$$

which accounted for 50% of the variance in one extensive set of data.[43] Because ΔL decreases with level instead of remaining constant as called for by Weber's law, this improvement in discrimination is called the near miss to Weber's law. The decrease in the size of ΔL with level also goes counter to the hypothesis that discrimination is a direct reflection of the slope of the loudness function, being better whenever the slope is steeper. In fact, discrimination is poorest near threshold, where the slope is steepest. Furthermore, a direct test of the hypothesis showed that level discrimination was the same for a tone heard in narrow-band noise as for an equally loud tone in broadband noise despite a large difference in the slopes of the partially masked loudness functions.[44]

At frequencies above 6 kHz or so, discrimination depends on level in more complex ways.[45,46] Between 8 and 12 kHz discrimination worsens up to moderate levels and then first begins to improve with further increases in level. At 14 kHz, discrimination changes little with level. Figure 7 plots ΔL as a function of tone level for eight studies over a range of frequencies from 0.25 to 14 kHz. The solid line from Florentine et al.[46] tends to be higher than results of many other studies, probably because their six listeners were not as well trained and motivated.

As implied above, level discrimination is independent of frequency only up to 6 kHz or so; it worsens (ΔL increases) with frequency starting at about 8 kHz, especially at levels greater than 50 dB SPL.[46] As discussed below, this deterioration can be ascribed to the restricted spread of excitation at high frequencies.

Duration Level discrimination improves as stimulus duration increases up to anywhere from 12 ms to 2 s, depending on the investigation and the sound frequency.[47] Reasons for the wide divergence among studies may be related to listener variables.

Bandwidth Level discrimination improves as a band of noise is widened. Unlike pure-tone discrimination, for a broadband stimulus such as white noise[48] or a widely spaced 18-tone complex,[49] discrimination improves with level only up to about 20 or 30 dB SPL. At higher levels ΔL is constant, meaning that Weber's law applies. Thus, near threshold, discrimination for broadband noise is no better than for a pure tone, at moderate levels discrimination is better, and above about 75 dB SPL it is worse.

Listener Variables Listeners differ markedly in their ability to discriminate between successive bursts. Florentine et al.[46] reported ratios of 2 : 1 and even 3 : 1 in the ΔL of one listener compared to another for the identical tone. Although Florentine et al.[46] were able to reduce the size of their listeners' ΔL's by giving them extensive practice and monetary reward, they did not report a concomitant reduction in intersubject variability.

Cochlear impairment has been thought to lead to improved level discrimination. However, the improvement is seen only when impaired listeners are compared to normally hearing listeners at the same SLs where the SPLs are higher for the impaired listeners. Compared at the same SPLs (lower SLs for the impaired), impaired listeners are no better and are often poorer than normally hearing listeners; just how much poorer depends on the shape of the impaired listener's threshold curve.[47]

3.2 Increment Detection

In the older literature, reference was often made to the measurements of Riesz,[50] who used sinusoidal amplitude modulation of continuous tones presented at many levels and frequencies. However, such measurements are contaminated by too many variables, such as adaptation, to give a clear picture of the ability to discriminate loudness differences.

Nonetheless, measurements of the minimum detectable increment in an ongoing sound are of importance in understanding the neural code for loudness. Measurements are available for tones at various frequencies and levels and for bands of noise. An important variable has been the time difference between the onset of the pedestal and the onset of the increment as well as such factors as the duration and rise–fall time of the increment. The interactions among these variables are complex, but a few general statements can be made.

Harris,[51] in an extensive series of measurements of increment detection, found that ΔL varied little with signal frequency. The increments lasted 150 ms and had a rise–fall time of 10 ms and occurred once a second. The effect of level was like that for burst comparison; ΔL decreased rapidly with level up to about 30 dB and more slowly thereafter. This presentation mode yields results more like that for burst comparison than for sinusoidal modulation. Harris also showed that the ΔL decreased as the duration of the increment increased from 20 to between 150 and 300 ms; the decrease was striking at sensation levels below 40 dB. At 40 dB and above, the ΔL decreased much less rapidly with duration. Duration would be expected to play less of a role in increment detection whenever onset is most important. Thus at low levels onset would seem to be less important than the excitation pattern evoked by the whole increment.

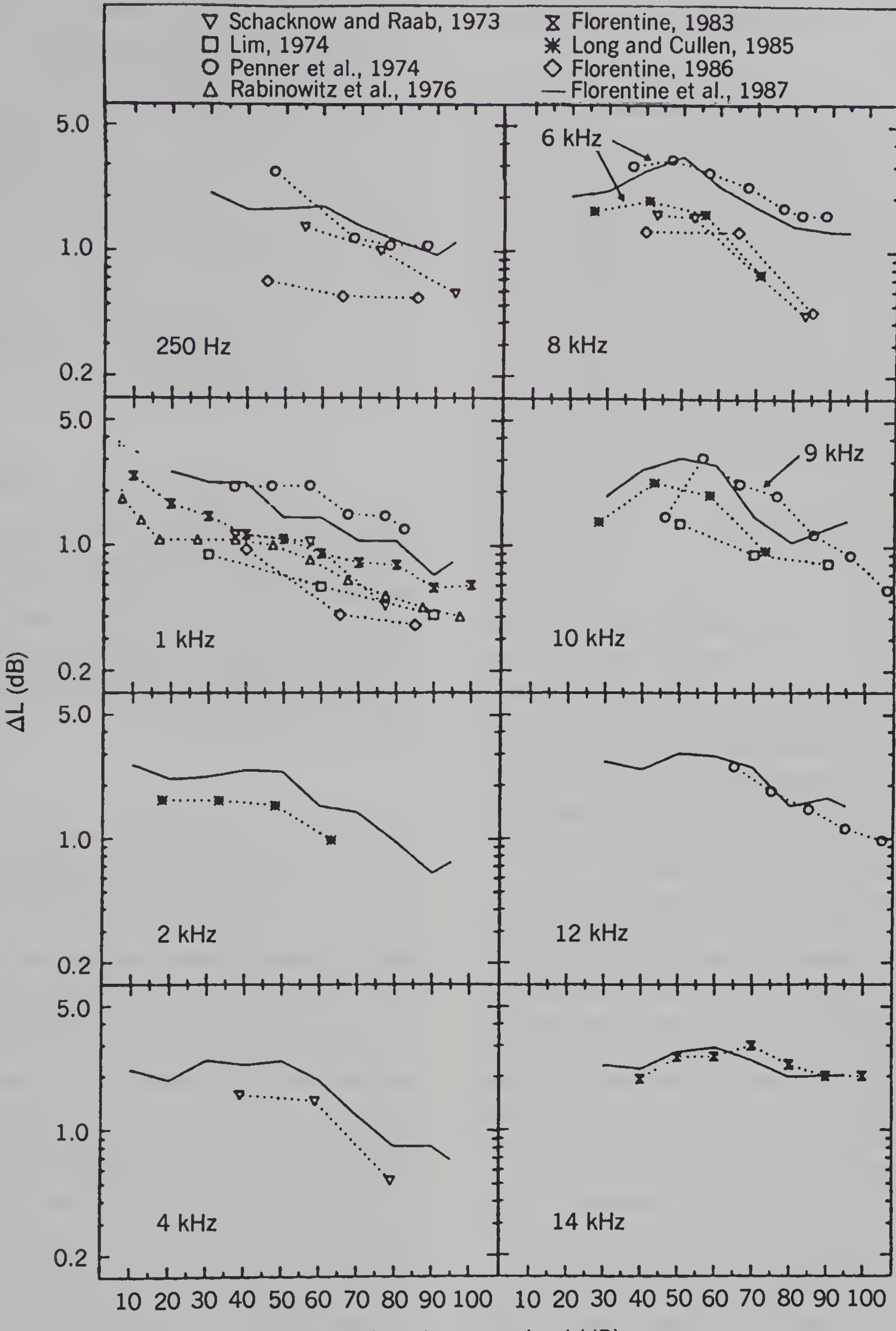


Fig. 7 Measured ΔL from eight studies. (Adapted from Ref. 46.)

The importance of the onset depends on frequency as well as level. Increment detection improves with increases in the delay between the onset of the base or pedestal sound and the onset of the increment;[52] the improvement is greater at 4 kHz than at 0.5 kHz and is greater at higher levels. Because detection continued to improve as the increment onset was delayed many seconds, some kind of long-term neural adaptation is likely to be involved; however, loudness adaptation is not directly implicated.

4 MODELS OF LOUDNESS AND LOUDNESS DISCRIMINATION

Although the neural code for loudness remains unclear, it is generally thought to be related to the quantity of neural activity in the auditory system. The more neurons that respond to a sound and the faster they respond, presumably the greater the loudness. Logical as this analysis may seem, the physiological data do not provide clear support and some even speak against it. The toughest problem has been to identify a physiological response capable of encoding the immense range of sound pressures to which organisms are sensitive, both with respect to loudness magnitude and discrimination; loudness increases from near inaudibility at threshold to intolerable levels over a range of sound pressures of more than a million to one. Yet, level discrimination remains extremely fine throughout. The demonstration by Liberman[53] that 10–15% of primary auditory nerve fibers have very high absolute thresholds, as much as 70 dB higher than the most sensitive fibers, provided another physiological basis for encoding intensity. These findings have also provided a basis for physiological models of intensity discrimination.[54] (Recent and comprehensive reviews of these matters are available.[55,56]) Nonetheless, because these physiological data have not yet led to quantitative models, the following discussion focuses on models that emphasize the quantity of assumed excitation in the auditory system.

Perhaps the most successful and best known of these psychophysical models are Zwicker's[35,57,58] for loudness and its extension by Florentine and Buus[49] to discrimination. The models are based on Fletcher and Munson's[59] proposal that masking patterns (such as revealed by measuring the threshold for a pure tone as a function of frequency in the presence of a narrow-band masker) reflect the level and spread of excitation in the auditory system. The excitation pattern thus derived determines loudness and provides the basis for loudness discrimination.

Hellman and Zwicker[60] provided an interesting, recent example of the model's application to loudness. They compared the measured and calculated loudness of sounds composed of broadband noise and a tonal component. Increasing the overall level of the complex sound by increasing the noise led to greater loudness, but increasing the overall level by increasing the level of the tone (for high tone-to-noise ratios of 20–30 dB re third-octave band of noise) and simultaneously decreasing the level of the noise diminished overall loudness. Figure 8 shows how the loudness of the complex increased with the overall level of tone plus noise. Also shown is the loudness calculated from Zwicker's model. (A convenient computer program for calculating loudness by this model is available.[61]) The points on the graph represent different tone-to-noise ratios, ranging from +5 to +30 dB (as well as the overall levels shown on the abscissa). Calculated and measured values are placed so as to reveal their close agreement with respect to the nonmonotonic relation between loudness and overall SPL; they differ mainly in the rate at which loudness increases with overall level. In the model, loudness is calculated by integrating over all the sound frequencies in the excitation pattern. When overall level is increased by taking a great deal of energy from the broadband noise and concentrating it at a single frequency—as happens at high tone-to-noise ratios—the height of the loudness pattern at all frequencies except those near the tonal component is reduced. Owing to the slow rate at which loudness changes with amplitude, the increase in height around the tonal component is too small to compensate for the lowered height at all the other frequencies.

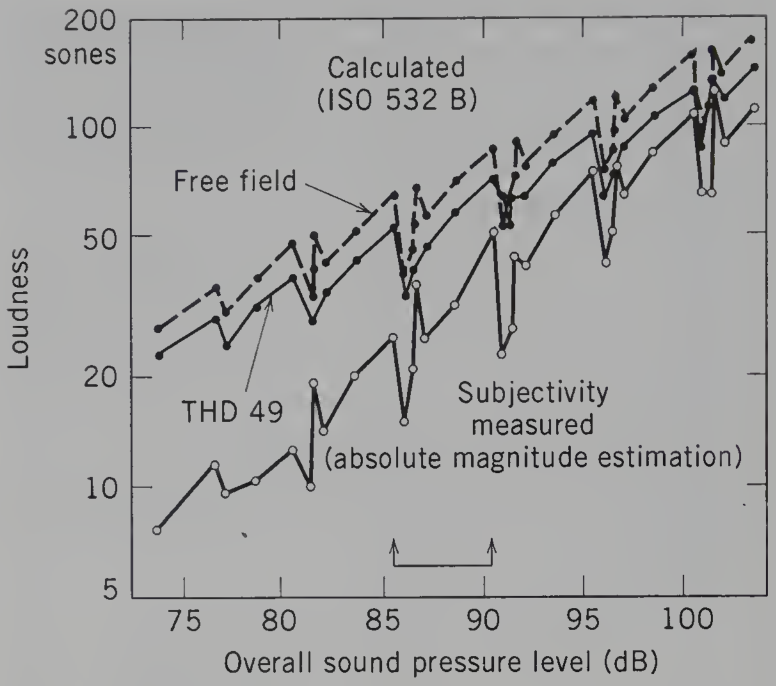


Fig. 8 Measured and calculated loudness functions for a tone–noise complex composed of 1-kHz tone and broadband noise. The dashed curve shows the calculated loudness function for listening in a free field, the dotted curve shows the calculated loudness function for listening with TDH-49 earphones, and the solid curve shows the measured loudness function. (From Ref. 60.)

The excitation pattern model applied by Florentine and Buus[49] to discrimination data helped explain many of the effects of level and frequency on discrimination. For example, the near miss to Weber's law arises from the spread of excitation as a pure tone is increased in level, providing the system with more and more channels for the comparison of one sound to another. A white noise does not show the near miss because it already stimulates much of the system (except close to threshold where, indeed, ΔL does decrease with increasing level).

Moreover, Weber's law can be recovered, as Viemeister[62] has shown, by partially masking a tone with a high-pass masker that eliminates the high-frequency channels and so prevents discrimination from improving with level, in perfect accord with the excitation pattern model. Very high frequencies also obey Weber's law, presumably because few channels remain into which their excitation can extend with increasing level. Despite the success of the excitation pattern models in handling many psychophysical data for both loudness and loudness discrimination, some data require other assumptions about the physiology underlying loudness[63] and about the possible role of more complex variables such as stimulus context.[64]

Acknowledgments

I thank Rhona Hellman and Georges Canévet for helpful comments on this manuscript. Wise advice from the editors, David M. Green and Dennis McFadden, was most useful. Support for the completion of this chapter was provided in part by a grant, RO1NS07270, from the National Institutes of Health.

REFERENCES

1. International Standardization Organization (ISO), "Method for Calculating Loudness Level," R532, ISO, New York, 1966.
2. B. Scharf and J. C. Stevens, "The Form of the Loudness Function near Threshold," *Proceedings of the Third International Congress on Acoustics*, Vol. 1, Elsevier, Amsterdam, 1961, pp. 80–82.
3. L. E. Marks and J. C. Stevens, "The Forms of the Psychophysical Function near Threshold," *Perception and Psychophysics*, Vol. 4, 1968, pp. 315–318.
4. J. J. Zwislocki, "Analysis of Some Auditory Characteristics," in R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol. III, Wiley, New York, 1965.
5. L. E. Humes and W. Jesteadt, "Models of the Effects of Threshold on Loudness Growth and Summation," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 1933–1943.
6. R. P. Hellman and C. H. Meiselman, "Loudness Relations for Individuals and Groups in Normal and Impaired Hearing," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 2596–2606.
7. G. Canévet, R. Hellman, and B. Scharf, "Group Estimation of Loudness in Sound Fields," *Acustica*, Vol. 60, 1986, pp. 277–282.
8. H. F. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement and Calculation," *J. Acoust. Soc. Am.*, Vol. 5, 1933, pp. 82–108.
9. D. W. Robinson and R. S. Dadson, "A Re-determination of the Equal-Loudness Relations for Pure Tones," *Br. J. Appl. Phys.*, Vol. 7, 1956, pp. 166–181.
10. International Standardization Organization (ISO), "Normal Equal-Loudness Contours for Pure Tones and Normal Threshold of Hearing under Free Field Listening Conditions," R226, ISO, New York, 1962.
11. K. Betke and R. Weber, "Re-examination of Equal Loudness Contours and Minimum Audible Field," *Proceedings of the Thirteenth International Congress on Acoustics*, Vol. I, Dragan Srnic, Belgrade, 1989, pp. 483–486.
12. R. P. Hellman and J. J. Zwislocki, "Loudness Determination at Low Sound Frequencies," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 60–64.
13. B. Schneider, A. A. Wright, W. A. Edelheit, P. Hock, and C. Humphrey, "Equal Loudness Contours Derived from Sensory Magnitude Judgments," *J. Acoust. Soc. Am.*, Vol. 51, 1972, pp. 1951–1959.
14. E. Zwicker, "Procedure for Calculating Loudness of Temporally Variable Sounds," *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 675–682.
15. B. Scharf, "Loudness of Complex Sounds as a Function of the Number of Components," *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 783–785
16. E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical Bandwidth in Loudness Summation," *J. Acoust. Soc. Am.*, Vol. 39, 1957, pp. 548–557.
17. B. Scharf, "Loudness Summation and Spectrum Shape," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 228–233.
18. B. Scharf and A. Houtsma, "Audition II," in K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol. 1, Wiley, New York, 1986.
19. O. J. Pedersen, P. E. Lyregaard, and T. E. Poulsen, "The Round Robin Test on Evaluation of Loudness Level of Impulsive Noise," Report No. 22, Acoustics Laboratory, Technical University of Denmark, Copenhagen, 1977.
20. A. Miskiewicz, B. Scharf, C. Meiselman, and R. Hellman, "Loudness Adaptation at High Frequencies," *J. Acoust. Soc. Am.*, Vol. 94, 1993, pp. 1281–1286.
21. B. Scharf, "Loudness Adaptation," in J. V. Tobias and E. D. Schuber (Eds.), *Hearing Research and Theory*, Vol. 2, Academic, New York, 1983.
22. G. Canévet, B. Scharf, and M.-C. Botte, "Simple and Induced Loudness Adaptation," *Audiology*, Vol. 24, 1985, pp. 430–436.
23. K. Gjaevenes and E. R. Rimstad, "The Influence of the Rise Time on Loudness," *J. Acoust. Soc. Am.*, Vol. 51, 1972, pp. 1233–1239.
24. S. Kuwano, S. Namba, H. Miura, and H. Tachibana, "Evaluation of the Loudness of Impulsive Sounds Using Sound Exposure Level Based on the Results of a Round Robin Test in Japan," *J. Acoust. Soc. Jpn.*, (E), Vol. 8, 1987, pp. 241–248.
25. E. Port, "Ueber die Lautstärke einzelner kurzer Schallimpulse," *Acustica*, Vol. 13, 1963, pp. 212–223.
26. Scharf, B. "Loudness," in E. C. Carterette and M. P. Fried-

man (Eds.), *Handbook of Perception*, Vol. 4, *Hearing*. Academic, New York, 1978, pp. 187–242.

27. T. Sone, et al., "Loudness and Noisiness of a Repeated Impact Sound: Results of Round Robin Tests in Japan (II)," *J. Acoust Soc. Jpn., (E)*, Vol. 8, 1987, pp. 249–261.
28. B. Berglund and S. Nordin, "Loudness and Brainstem Auditory Evoked Response for Intermittent Sound," *Proceedings of the Thirteenth International Congress on Acoustics*, Vol. I, Dragan Srnic, Belgrade, 1989, pp. 277–380.
29. S. Namba, "Loudness Varying with Time," *Proceedings of the Thirteenth International Congress on Acoustics*, Vol. I, Dragan Srnic, Belgrade, 1989, pp. 361–364.
30. S. Kuwano and H. Fastl, "Loudness Evaluation of Various Kinds of Non-Steady State Sound Using the Method of Continuous Judgment by Category," *Proceedings of the Thirteenth International Congress on Acoustics*, Vol. I, Dragan Srnic, Belgrade, 1989, pp. 365–368.
31. G. Canévet and B. Scharf, "The Loudness of Sounds That Increase or Decrease Continuously in Level," *J. Acoust Soc. Am.*, Vol. 88, 1990, pp. 2136–2142.
32. R. S. Schlauch, "A Cognitive Influence on the Loudness of Tones That Change Continuously in Level," *J. Acoust Soc. Am.*, Vol. 92, 1992, pp. 758–765.
33. S. S. Stevens and M. Guirao, "Loudness Functions under Inhibition," *Percept. Psychophys.*, Vol. 2, 1967, pp. 459–465.
34. R. P. Hellman, "Effect of Noise Bandwidth on the Loudness of a 1000 Hz Tone," *J. Acoust Soc. Am.*, Vol. 48, 1970, pp. 500–504.
35. E. Zwicker, "Ueber die Lautheit von ungedrosselten und gedrosselten Schallen" *Acustica*, Vol. 13 (Beiheft 1), 1963, pp. 194-211.
36. B. Scharf, "Partial Masking," *Acustica*, Vol. 14, 1964, pp. 16–23.
37. L. E. Marks, "Sensory and Cognitive Factors in Judgments of Loudness," *J. Exper. Psychol.: Human Percept. Perform.*, Vol. 5, 1979, pp. 426–443.
38. W. Riach, D. N. Elliott, and J. C. Reed, "Growth of Loudness and Its Relationship to Intensity Discrimination under Various Levels of Auditory Fatigue," *J. Acoust Soc. Am.*, Vol. 34, 1962, pp. 1764–1767.
39. M.-C. Botte and S. Monikheim, "New Data on Short-Term Effects of Tone Exposure," *J. Acoust Soc. Am.*, Vol. 95, 1994, pp. 2598–2605.
40. M. Florentine, S. Buus, B. Scharf, and E. Zwicker, "Frequency Selectivity in Normally-Hearing and Hearing-Impaired Observers," *J. Speech Hear. Res.*, Vol. 23, 1980, pp. 646–669.
41. B. Scharf and R. P. Hellman, "A Model of Loudness Summation Applied to Impaired Ears," *J. Acoust Soc. Am.*, Vol. 40, 1966, pp. 71–78.
42. W. M. Rabinowitz, J. S. Lim, L. D. Braida, and N. I. Durlach, "Intensity Perception. VI. Summary of Recent Data on Deviations from Weber's Law for 1000-Hz Tone Pulses," *J. Acoust Soc. Am.*, Vol. 59, 1976, pp. 1506–1509.
43. W. Jesteadt, C. C. Wier, and D. M. Green, "Intensity Discrimination as a Function of Frequency and Sensation Level," *J. Acoust Soc. Am.*, Vol. 61, 1977, pp. 169–177.
44. R. Hellman, B. Scharf, M. Teghtsoonian, and R. Teghtsooinian, "On the Relation between the Growth of Loudness and the Discrimination of Intensity for Pure Tones," *J. Acoust Soc. Am.*, Vol. 82, 1987, pp. 448–453.
45. R. P. Carlyon and B. C. J. Moore, "Intensity Discrimination: A Severe Departure from Weber's law," *J. Acoust Soc. Am.*, Vol. 76, 1984, pp. 1369–1376.
46. M. Florentine, S. Buus, and C. R. Mason, "Level Discrimination as a Function of Level for Tones from 0.25 to 16 kHz," *J. Acoust Soc. Am.*, Vol. 81, 1987, pp. 1528–1541.
47. B. Scharf and S. Buus, "Audition I," in K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of Perception and Human Performance*, Vol. 1, Wiley, New York, 1986.
48. A. J. Houtsma, N. I. Durlach, and L. D. Braida, "Intensity Perception XI. Experimental Results on the Relation of Intensity Resolution to Loudness Matching," *J. Acoust Soc. Am.*, Vol. 68, 1980, pp. 807–813.
49. M. Florentine and S. Buus, "An Excitation-Pattern Model for Intensity Discrimination," *J. Acoust Soc. Am.*, Vol. 70, 1981, pp. 1646–1654.
50. R. R. Riesz, "Differential Intensity Sensitivity of the Ear," *Phys. Rev.*, Vol. 31, 1928, pp. 867–875.
51. J. D. Harris, "Loudness Discrimination," *J. Speech Hear. Disord.*, Vol. 11, 1963, pp. 1–63.
52. B. Scharf, G. Canévet, and L. M. Ward, "On the Relation between Intensity Discrimination and Adaptation," in Y. Cazals, L. Demany, and K. Horner (Eds.), *Auditory Physiology and Perception*, Pergamon, Oxford, 1992, pp. 289–295.
53. M. C. Liberman, "Auditory-Nerve Response from Cats Raised in a Low-Noise Chamber," *J. Acoust Soc. Am.*, Vol. 63, 1978, pp. 442–455.
54. R. L. Winslow and M. B. Sachs, "Effect of Electrical Stimulation of the Crossed Olivocochlear Bundle on Auditory Nerve Response to Tones in Noise," *J. Neurophysiol.*, Vol. 75, 1987, pp. 1002–1021.
55. N. F. Viemeister, "Psychophysical Aspects of Auditory Intensity Coding," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function: Neurobiological Bases of Hearing*, Wiley, New York, 1988, pp. 213–241.
56. R. L. Smith, "Encoding of Sound Intensity by Auditory Neurons," in G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), *Auditory Function: Neurobiological Bases of Hearing*, Wiley, New York, 1988, pp. 243–274.
57. E. Zwicker and B. Scharf, "A Model of Loudness Summation," *Psychol. Rev.*, Vol. 72, 1965, pp. 3–26.
58. B. C. J. Moore and B. R. Glasberg, "The Role of Frequency Selectivity in the Perception of Loudness, Pitch and Time," in B. C. J. Moore (Ed.), *Frequency Selectivity in Hearing*, Academic, London, 1986, pp. 251–308.

59. H. F. Fletcher and W. A. Munson, "Relation between Loudness and Masking," *J. Acoust Soc. Am.*, Vol. 9, 1937, pp. 1–10.
60. R. Hellman and E. Zwicker, "Why Can a Decrease in dB(A) Produce an Increase in Loudness?" *J. Acoust Soc. Am.*, Vol. 82, 1987, pp. 1700–1705.
61. E. Zwicker, H. Fastl, and C. Dallmayr, "Basic Program for Calculating the Loudness of Sounds from Their 1/3-oct-Band Spectra According to ISO 532B," *Acustica*, Vol. 55, 1984, pp. 63–67.
62. N. F. Viemeister, "Intensity Discrimination of Pulsed Sinusoids: The Effects of Filtered Noise," *J. Acoust Soc. Am.*, Vol. 51, 1972, pp. 1265–1269.
63. F-G. Zeng and C. W. Turner, "Binaural Loudness Matches in Unilaterally-Impaired Listeners," *Quart. J. Exper. Psychol.*, Vol. 43A, 1991, pp. 565–583.
64. L. E. Marks, "Recalibrating the Auditory System: The Perception of Loudness," *Percept. Psychophys.*, Vol. 20, 1994, pp. 382–396.

119

EFFECTS OF HIGH-INTENSITY SOUND

W. DIXON WARD

1 INTRODUCTION

Sounds of high intensity make speech communication difficult, if not impossible, and may render warning signals inaudible. They elicit changes in the autonomic system, such as dilation of the pupil of the eye, increased heart rate, and vasoconstriction of the extremities and, if their onset is sudden, evoke the startle response. They may degrade performance of certain tasks, interfere with sleep, and induce irritation ranging from annoyance to rage. At very high levels, sound becomes painful and may even directly affect the function of the visual and vestibular systems as well as of internal organs.

The main aftereffect of high-intensity sounds, however, is on the auditory system, manifested principally in the form of a temporary or permanent loss of sensitivity and acuity, although tinnitus and suppression of otoacoustic emissions may also occur. Permanent hearing loss can be caused by short exposures to extremely high sound levels or by repeated exposures to more moderate levels; the extent of the damage produced in a particular ear is determined not only by the duration and level of the exposure but also by its spectrum and temporal pattern as well as by individual characteristics of the auditory system concerned.

Exposures that result in permanent hearing loss can be regarded as stressors that may produce other aftereffects, particularly on the circulatory system. However, there is little evidence that other stressors such as vibration, heat, or exercise act synergistically with high-intensity sound to produce hearing loss.

Until a few years ago, an "intense" sound was commonly understood to mean one that overloaded the hearing mechanism, that is, an acoustic stimulus that elicited a nonlinear response from the auditory system. However, with the discovery of the essential nonlinearity of the cochlea whereby energy released at the basilar membrane enhances the vibratory motion of the cochlear partition at the lowest values of input energy, it has become clear that the auditory system as a whole is never really linear, so some other definition of *intense* must be found. A suitable alternative is suggested by a different characteristic of the auditory system, namely, a sound capable of arousing the acoustic reflex. Whether the stiffening of the ossicular chain produced by this reflex, which reduces the intensity of the sound reaching the inner ear, is regarded as providing protection against one's own voice or against other sounds, the auditory system apparently judges some sounds to be "too loud" and takes steps to reduce them.

2 CONCOMITANT EFFECTS

2.1 Middle-Ear Muscles

If the foregoing definition of intense sound is adopted, then one of its primary effects is, circularly, contraction of the stapedius and tensor tympani muscles. Because the strength of their contraction depends on the loudness of the sound rather than on its intensity, all of the physical characteristics of sound that determine loudness are relevant: not only overall intensity but also frequency, duration, and spectral width. However, for sounds of more than half a second in duration and in the audible range, strength of contraction increases with intensity, beginning at a threshold value of 85–90 dB sound pressure level (SPL) for single pure tones, 70–75 dB SPL

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

for broadband noise, and reaching a maximum at levels about 30 dB higher in either case.[1]

Although the latency of the first muscular response to an arousal sound is only 10–15 ms, full contractile strength is not reached until as much as 150 ms after onset.[2] In the case of a sustained arousal sound, the muscles do not maintain full strength indefinitely, but relax (adapt) more or less rapidly depending on the frequency and temporal structure of the arouser. The lower the frequency of the sound, or the greater its fluctuation in time, the more slowly will the muscles adapt. As the increased muscular tension produces primarily an increase in the stiffness of the middle-ear mechanism, low frequencies (below 1000 Hz) are attenuated more than high frequencies by the reflex.

Other mechanisms do exist that may modify the sound reaching the inner ear. As sound level is increased, peak clipping eventually will occur when the motion of the eardrum or the stapes reaches a limit imposed by the annulus or annular ligament, respectively.[3] This apparently happens only for levels above 120–130 dB, at least in the anesthetized cat.[4] In addition, Bekesy's observations on human cadaver ears[5] revealed that at a similar level the axis of rotation of the stapes may suddenly shift by 90°, thereby reducing the sound transmitted to the cochlea. In both of these cases, however, the middle-ear muscles were inoperative, so just how all of these three influences interact in the normal ear is not known. One can expect, however, that some sort of discontinuity of function may occur at around 120 dB SPL.

2.2 Masking

High-intensity sounds may render other sounds inaudible. In the case of direct masking—that is, when the masking sound (the masker) is in the same frequency region as the sound that is masked (the signal)—the intensity of the masker is irrelevant, in the sense that the masked threshold of the signal will be linearly related to the level of the masker, whether the latter is weak or intense: An increase of *X* decibels in masker level will produce an *X*-decibel increase in the threshold of the signal. If the signal is speech, then its intelligibility will remain essentially constant if it were at the same time increased by that same *X* decibels.

The foregoing generalization is not true when the signal and masker frequencies f_s and f_m are different. In either upward spread of masking ($f_m < f_s$) or remote masking ($f_m > f_s$), the rate of growth with masker level increases with level, as the distortion products—aural harmonics, combination tones, and perhaps envelope detection—generated in the cochlea by the masker also grow.[6] So by the time the masking level reaches the intense range, considerable upward spread of masking already exists for any type of masker—pure tone, tonal complex, or narrow band of noise. Fig. 1 shows the masking pattern developed by a 50-Hz pure tone at 100, 115, and 130 dB SPL[7]; at the highest level, the pattern of excitation on the basilar membrane apparently covers the entire auditory range.

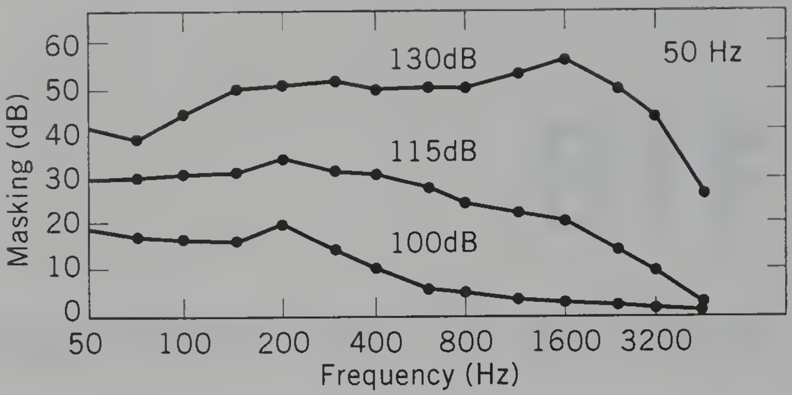


Fig. 1 Upward spread of masking produced by a 50-Hz pure tone at high intensities (100, 115, and 130 dB SPL). Note that the growth of masking is highly nonlinear, increasing at 1.6 kHz by more than 30 dB as the masking level is raised from 115 to 130 dB SPL. (From Ref. 7.)

Remote masking by sounds above 100 dB has not been extensively studied beyond the original observations of Bilger and Hirsh.[6] Figure 2 shows their results for a masker that was a half-octave band of noise and illustrates the characteristics of spread of masking mentioned above.

The spread of cochlear activity at high intensities probably accounts for the fact that the masking of speech

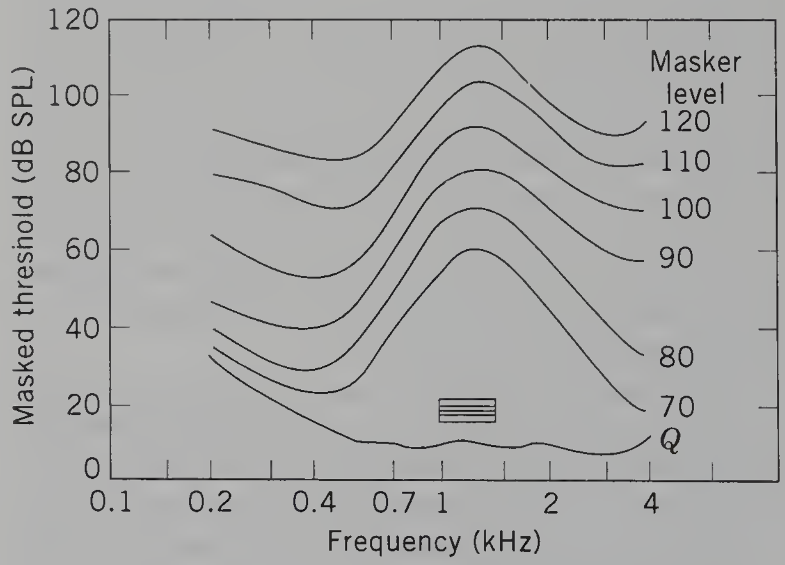


Fig. 2 Growth of remote, direct, and upward spread of masking with level of masker. The ordinate indicates the absolute thresholds at various frequencies (abscissa) as a function of masker level in decibels of SPL (parameter). Thresholds in quiet are represented by *Q*. The dark bar indicates the frequency range of the half-octave masker (1000–1420 Hz). (After Ref. 6.)

produced by loud noises depends very little on the precise spectral characteristics of the sound unless it is a pure tone. Without amplification, speech cannot be understood in any typical 90-dB(A) noise, even if the talker is shouting, unless the listener's ear is only a foot or so from the talker.[8] Even with amplification to keep the overall speech-to-noise ratio constant, speech intelligibility declines once the level exceeds 100 dB(A), dropping rapidly above 120 dB(A).[9]

2.3 Temporal Changes in Masking

A difference exists between types of masking in terms of the stability of masking by a sustained high-intensity sound. Direct masking shows essentially no change in masked threshold of a tone with time; remote masking, both ipsilateral and contralateral, displays gradual improvement (less masking) during a period of several minutes; however, speech intelligibility appears to deteriorate over at least the first 15 min of exposure to a broadband noise at 125–130 dB SPL.[9] The reason for these differences is not completely clear. Logically, direct masking would be expected to be invariant with time, since any time-linked process that would affect the noise should also affect the signal equally. The gradual loss of intelligibility of speech may be associated with the development of pronounced auditory fatigue. However, the decrease of remote masking with time has no simple explanation. A gradual relaxation of the middle-ear-muscle reflex, allowing an increasing amount of energy at the low frequency of the test signal to enter the cochlea, would produce an effect in the indicated direction; however, a change in remote masking is observed in auditory systems in which the middle-ear muscles are either excised[10] or paralyzed.[11] Furthermore, this change in remote masking occurs with sounds that are not intense enough to arouse the middle-ear muscles. Thus remote masking may be primarily caused by the cochlear excitation associated with distortion products developed in the ear, and its change with time should reflect some change in these distortion products, the nature of which is still obscure. However, the efferent system that may produce changes in the mechanical characteristics of the organ of Corti with sustained stimulation[12] is probably involved.

2.4 Physiological Reactions

Sudden intense sounds will evoke the startle response: an immediate contraction of the orbital eye muscles and the flexor muscles of the legs, arms, and back, manifested as an eyeblink and a crouching movement.[13] This startle response may be followed by an orienting reflex in which the head and eyes turn toward the sound source. The magnitude of the startle response is related directly to the loudness of the sound and inversely to its rise time but is also affected by its unexpectedness and by the level of the background ambient noise.[14,15] Davis et al.[14] found that for 2-s bursts of a 1-kHz tone the muscular response increased dramatically between levels of 90 and 120 dB. Repetition of the sound will usually result in a reduction of the response, although such habituation may not become complete. Even when no overt movement can be seen, transient muscular activity detected by electrophysiological techniques may interfere with an ongoing motor task involving those muscles. This lack of complete habituation, especially to gunfire involving peak levels of 140 dB or over, is probably responsible for the successful promotion of the use of hearing protectors at shooting ranges. Shooters accept earplugs and muffs not so much for protection against hearing loss as for an improvement in accuracy that occurs when the reports from other shooters' guns are attenuated enough that they no longer produce an involuntary flinch.

If the intense sound continues beyond a few minutes, other physiological changes will be manifested. Some of these are clearly linked to the startle response in the sense that the change was initiated by the abrupt onset of the sound but could be measured only several seconds later; these include a short-lived increase in blood pressure and heart rate and a decrease in rate of breathing.[14] Other changes, however, are more persistent. Of these, the two most easily measured and hence most extensively studied are a vasoconstriction of the extremities, indicated by measurements of finger volume, and a dilation of the pupil of the eye.[16] Others include decreased skin resistance and various biochemical and endocrinological reactions. Most of these are among those included in the "stress reaction," so their occurrence indicates that high-intensity sound can be considered to be a stressor. However, the magnitude of the observed effect (changes in characteristics of the circulatory systems, including blood components such as leukocytes, eosinophils, and basophils; composition of urinary excretions of endocrinic substances such as epinephrine, norepinephrine, and other corticoids) associated with sound exposures that do not cause hearing loss lie within the range of normal variability of these indicators in everyday life.[17] So it is not surprising that, despite innumerable studies of these stress-induced reactions, it has still not been shown convincingly that sound alone can produce any permanent aftereffects on the autonomic or circulatory systems.

2.5 Tactile Effects

At high levels, sounds will produce tactile sensations in the middle ear, the nature of which is relatively independent of the spectral characteristics (there are no tactile receptors in the inner ear). Early systematic studies of this phenomenon indicated that "discomfort" usually began at about 120 dB SPL, "tickle" at 130 dB SPL, and "pain" at 140 dB SPL.[18] However, individual differences in discomfort are large, with some persons reporting sounds to be "uncomfortably loud" at 80 or 90 dB SPL, although the average normal listener needs about 105 dB.[19] Totally deaf individuals without eardrums reported no pain even when exposed to jet noise at 170 dB SPL,[20] so the eardrum may be the primary source of pain.

Explosions may rupture the eardrum and/or disrupt the ossicular chain. Such events do not occur under controlled conditions in humans, so the sound levels involved have not been measured; however, the evidence indicates that peak levels under 170 dB SPL will not cause this form of damage.

2.6 Vestibular Effects

Because the vestibular sense organ (the semicircular canals) is connected to the cochlea, pressure transmitted to the cochlea will also reach the semicircular canals. High intensities of sound might therefore be expected to affect the sense of balance. However, nystagmus, vertigo, and loss of balance occur only in levels of 140 dB SPL or greater[21] as long as the stimulation is bilaterally symmetrical. If only one ear is exposed, on the other hand, balance can be affected by sound levels of 100 dB or lower.[22]

2.7 Annoyance

Any sound made by others may generate annoyance in a particular individual. The magnitude of that annoyance does depend to some extent on the level of the sound: On average, the higher the level, the greater the interference with speech communication, mental concentration, task performance, recreational activities, and sleep and hence the greater the annoyance. However, there are so many other factors that determine the annoyingness of a sound (its spectrum, duration, and temporal pattern; the time of day; perceived avoidability, expectedness, meaningfulness, inter alia) that the role of level per se remains expressible only in qualitative terms. Indeed, widespread agreement does not exist on how to quantify the degree of annoyance, so that it is expressed in terms of the fraction of a population that is "highly annoyed".[23] It seems safe to say that any sound regarded as intense according to the present definition (90 dB SPL or higher for pure tones, 75 dB or above for noises) will become "highly annoying" to the average person if it is sufficiently prolonged.[24]

2.8 Miscellaneous Concomitant Effects

When one is exposed to the noise from an immobilized jet aircraft (generating 140 dB SPL near the exhaust from the engine or 160 dB SPL if the afterburner is operating), even with hearing protectors, numerous effects associated with the direct action of high-intensity low-frequency components of the noise on the body become apparent. Vision may be blurred because of vibration of the eyeballs, the teeth may ache, the sinuses "buzz," and the chest cavity and viscera are clearly being shaken at different rates. However, these intersensory effects occur only in sound levels that are so high that a few minutes of exposure would produce irreversible damage to the auditory system, so they have not been studied systematically.

3 AFTEREFFECTS OF INTENSE SOUND

The concomitant effects of high-intensity sounds are largely independent of their duration. This is not true of aftereffects, however. Duration of exposure is always an essential parameter, especially in the case of production of hearing loss. Although there is a persistant popular belief that there is an intensity of sound above which permanent damage to the auditory system will result, regardless of the duration of the sound, no empirical support for this belief exists: Any level that is "critical" in this sense is critical only for a particular duration.[25] On the other hand, duration and level are not the only parameters that must be specified in establishing a relation between a given exposure to high-intensity sound and its aftereffects. The spectral distribution and temporal pattern of the sound are also important. Thus the adoption by the International Standardization Organization (ISO) of the definition of "exposure" (E) as the integral of the square of the A-weighted sound pressure over the time concerned,

$$E = \int_{t_1}^{t_2} p_A^2(t)\, dt,$$

represents a drastic oversimplification of a complex issue.[26]

The problem here is that this definition embodies three assumptions: two that are demonstrably false—that spectra are equalized in effectiveness by the use of A weighting and that temporal pattern is irrelevant—and one assumption that is true only for a restricted set of conditions—that the total energy (the product of intensity and time, It, since I is proportional to p^2) involved in an exposure is the sole determining factor in predicting its effect. For this reason, the ISO definition of exposure will be rejected here. The term *exposure* will retain its original meaning: the entire history of pressure during a specified time period measured at the position of the listener's ear but with the listener absent.

3.1 Temporary Threshold Shift

The most well-known aftereffect of exposure to high-intensity sound is the change in auditory sensitivity. If the auditory threshold is measured before and after an exposure, the difference in hearing threshold levels (HTLs) is by definition the threshold shift (TS). If the shift later completely disappears, then it was a temporary threshold shift (TTS); if not, the final shift is termed a permanent threshold shift (PTS), and the initial shift was a compound threshold shift (CTS), that is, a combination of TTS and PTS.

It would be desirable to categorize threshold shifts in terms of the underlying physiological changes. For example, all of the following could conceivably be involved in TTS: (1) residual middle-ear-muscle activity; (2) displacement of the tectorial membrane relative to the basilar membrane; (3) changes in the chemical environment of the hair cells; (4) swelling of the hair cells, making stimulation more difficult in a mechanical sense; (5) an increase in internal noise, as for example due to increased blood flow ("pounding in the ears") or an audible tinnitus; (6) changes in or results of efferent activity at the basilar membrane; and of course (7) ordinary poststimulatory decrease in nerve excitability that could occur in the eighth nerve, cochlear nucleus, lateral lemniscus, inferior colliculus, medial geniculate, or acoustic cortex. However, until more knowledge about the relative contribution of these factors is available, classification of TTSs may be based on the duration of their combined effect: less than 1 s, 1 s to 2 min, 2 min to 16 h, and greater than 16 h.[27]

A TTS that lasts less than 1 s has been called *residual masking* or *forward masking*, although there is no reason to think that masking is actually involved (see Chapter 120). The magnitude of this ultra-short-term TTS, measured immediately after exposure to a short one, is proportional to the SPL of the exposure but relatively independent of its duration. The recovery process is exponential, in the sense that the TTS in decibels decreases linearly with the logarithm of recovery time.[28]

A TTS lasting a minute or so will follow a short exposure (a few minutes) at levels up to 80 dB SPL but is nearly independent of that level; that is, the magnitude of the shift is the same following exposure at 30 dB as at 80 dB. This variety of TTS, *low-level TTS*, is critically dependent on the procedure for determining threshold; it disappears rapidly if an interrupted test tone is used.

Intense sounds, if sustained long enough, will produce "ordinary" TTS, also called *physiological fatigue*: a shift that requires up to 16 h to disappear, with a course of recovery that, as in the case of forward masking, is linear in the logarithm of time (recovers exponentially).

The first minute or so following intense sound generally involves a "rushing noise" tinnitus that operates in the opposite direction of an objective noise. When the noise seems loudest, the threshold is lowest, and when the noise disappears (leaving a striking silence), the threshold is elevated. Recent work on this paradoxical noise indicates that it does not originate only in the cochlea but may involve higher centers of the auditory chain.[29]

Finally, TTSs that persist for more than 16 h, sometimes termed *pathological fatigue*, are produced by more severe exposures. This TTS recovers linearly in time (i.e, a certain number of decibels per day) rather than in log time, suggesting that the TTS in this case has a different physiological correlate. As much as 3 weeks can be required for this "delayed recovery" to be complete.[30]

In all cases, individual differences in the magnitude of the TTS caused by a particular exposure are huge. The problem of individual variability and the prediction of susceptibility will be discussed later. However, because of the possibility that TTS produced by a moderate exposure to a given sound spectrum might be correlated with the PTS resulting from a subsequent more severe exposure (longer, more intense, or given repeatedly), extensive study of ordinary TTS has provided knowledge about the exposure parameters involved in its generation.

TTS from Continuous Exposures Ordinary TTS is customarily measured 2 min after termination of exposure, that is, after the shorter lived recovery processes, which may produce a diphasic recovery curve, have run their course. For a continuous steady sound, this TTS_2 grows linearly with the logarithm of the exposure duration until it reaches an asymptote [asymptotic threshold shift (ATS)] after about 8 h of exposure. Also, TTS_2 grows linearly with sound level once the level has exceeded a base value below which only short-duration effects are produced, no matter how long the exposure. This base SPL is termed *effective quiet* (EQ), and the difference between EQ and a given SPL may be called the *effective level* (EL). EQ is about 85 dB SPL for pink noise and 70–80 dB for sounds an octave or narrower in bandwidth[31]; the exact value mirrors the transfer function of the outer and middle ears, being lowest for sounds in the 2–4-kHz region, where the resonance of the outer ear canal provides a 10–15-dB amplification of the measured field SPLs. For the typical industrial noise, EQ will probably have a value of 80 dB SPL; the ATS_2 produced by 8 h exposure at 100 dB SPL (EL = 100 − 80 = 20 dB) will be twice that produced by one at 90 dB (EL = 10 dB) and half of that produced by 120 dB SPL (EL = 40 dB). Since ATS grows at the rate of 1.7 dB per decibel of EL[32], the typical industrial noise will produce an ATS of 17 dB at 90 dB and 34 dB at 100 dB.

TTS from Varying Levels Prediction of the TTS produced by time-varying and intermittent noise exposures is somewhat complicated, because of partial recovery of hearing during the quieter periods. However, for relatively rapid interruptions or fluctuations (periods at a given level of 5 min or less), the TTS is proportional to the average effective level during the exposure, that is,

$$TTS_2 \propto \frac{1}{T} \int_0^t EL(t)\, dt.$$

Thus for the typical industrial noise, the same TTS_2 will be produced by 8 h exposure to a 90-dB steady noise, a 100-dB noise that is on half the time and off half the time, or a noise that changes at regular intervals from 80 to 90 to 100 dB and back to 80, or a noise that fluctuates irregularly but for which the average number of decibels by which 80 dB is exceeded is 10 dB.[33]

Temporary threshold shifts produced by patterns of longer exposures and rest periods can be calculated by means of empirical equations describing alternate growth and recovery of TTS_2. In this case also, however, the TTS is considerably less than the TTS that would have been produced by a steady noise over the same time but whose level was such as to keep the total energy constant [$L_{eq}(t)$]. For example, in the case of the noise that is on half the time at 100 dB and off the other half, L_{eq} is 97 dB, yet it produces only the same TTS as a 90-dB noise. Thus TTS is not a consistent function of the total energy of an exposure; how the energy is distributed in time makes a considerable difference in the magnitude of the TTS produced.

TTS from Impulse Noise The same is true for TTS from impulse noise, such as gunfire. Although 30 impulses of simulated gunfire at 150 dB SPL peak level may produce a TTS_2 of 20 dB in a particular ear, 300 impulses of the same pulse shape but at 140 dB SPL (i.e., the same total energy) will usually produce no TTS_2 in this individual.[34] Here it appears that there is a critical level (which, however, is a function of pulse duration) below which no effect is produced and above which the TTS increases with level. Unlike the growth of TTS_2 with time for steady or interrupted noises, the TTS_2 from impulses appears to be proportional to the number of impulses, as if each impulse produced the same amount of TTS in decibels.[35] For this reason, no attempts have been able to determine the point at which the TTS from repeated impulse noise reaches an asymptotic value.

Delayed Recovery The foregoing discussion assumes that TTS_2 does not exceed 40 dB; if this value is reached, or if even somewhat lower values of TTS_2 are the result of prolonged intermittent exposure to intensities of 110 dB or higher, recovery may shift to the "delayed" mode, in which little recovery occurs during the first hours after exposure and more than 16 h is required for full recovery.[36] It has long been suspected that delayed recovery indicates damage that is more likely to become irreversible if the next exposure occurs before full recovery,[37] although no substantial supporting evidence has been advanced.

3.2 Permanent Threshold Shift

The most undesirable aftereffect of exposure to high-intensity sound is PTS. Sound-induced PTS is commonly divided into two categories depending on whether the loss was produced by a single short exposure at a very high level (acoustic trauma) or by repeated longer exposures at more moderate levels. It is clear from animal studies that in acoustic trauma the inner ear has been subjected to such stress that its elastic limit, so to speak, has been exceeded. Various structures of the organ of Corti, including the hair cells, may become partly or wholly detached, and one or more of the several membranes in the cochlea may be ruptured, allowing intermixture of fluids of different composition, thereby poisoning the hair cells that survived the mechanical stress. The end result is a pronounced loss of sensitivity at the frequencies correlated with the locus of this destruction.

The physiological correlates of gradually induced PTS are not as clear. A fairly large PTS may exist despite a near-normal appearance of the cochlear structures, although the locus of what damage does exist will generally correspond to the frequencies showing the greatest PTS. Clearly, subtle defects in the morphological, chemical, and neurological characteristics of the cochlea are involved; these have been summarized recently by Saunders et al.[38]

Acoustic Trauma Little is known about acoustic trauma in humans, although it is not at all rare. Victims of acoustic trauma seldom have had a recent audiogram that would enable the amount of PTS to be determined with certainty. Even then, the exposure level and duration can usually only be grossly estimated. Finally, differences among people in susceptibility to damage are so great that single cases show that acoustic trauma is possible from a given exposure but not that it is inevitable or even likely.

The difference between likelihood of safety and possibility of damage is illustrated in Fig. 3. The circled symbols indicate the most severe single exposures that have been administered under controlled conditions to a group of at least five normal-hearing young men but that produced no PTS, while the symbols in squares show the estimated level and duration of exposures that apparently produced a PTS of 15 dB or more in at least one single individual. The probably safe exposures were all to broadband noise. The exposures included in the figure are from the classic studies of Davis et al.[39] on the effect of noise: point *D*, 32 min at 130 dB; point *E*, 1 min at 135 dB[40]; and point *SN*, 0.4 s at 153 dB, an exposure designed to be equivalent to the sound produced by the opening of an air bag.[41]

By contrast, a PTS of 15 min or more followed 1 min exposure to a 2-kHz tone at 130 dB (point *M*) or 8 min exposure to a 4-kHz tone at 120 dB SPL (point *S*) in one of several men so exposed in the Davis et al. study.[39] Davis himself received a 30-dB increase in his preexisting high-frequency hearing loss after exposure for 20 min to a 500-Hz tone at 140 dB SPL (point *H*). Point *O* indicates the effect of the ring of a cordless telephone in which the same transducer was used for the ringer and for voice, an exposure that has produced a measurable PTS in some small but unknown fraction of those individuals exposed,[42] and point *L* is an exposure of "a few seconds" to a tone of around 138 dB SPL that was being used routinely in an attempt to elicit the acoustic reflex, one that produced additional damage in two individuals who already had considerable loss.[43]

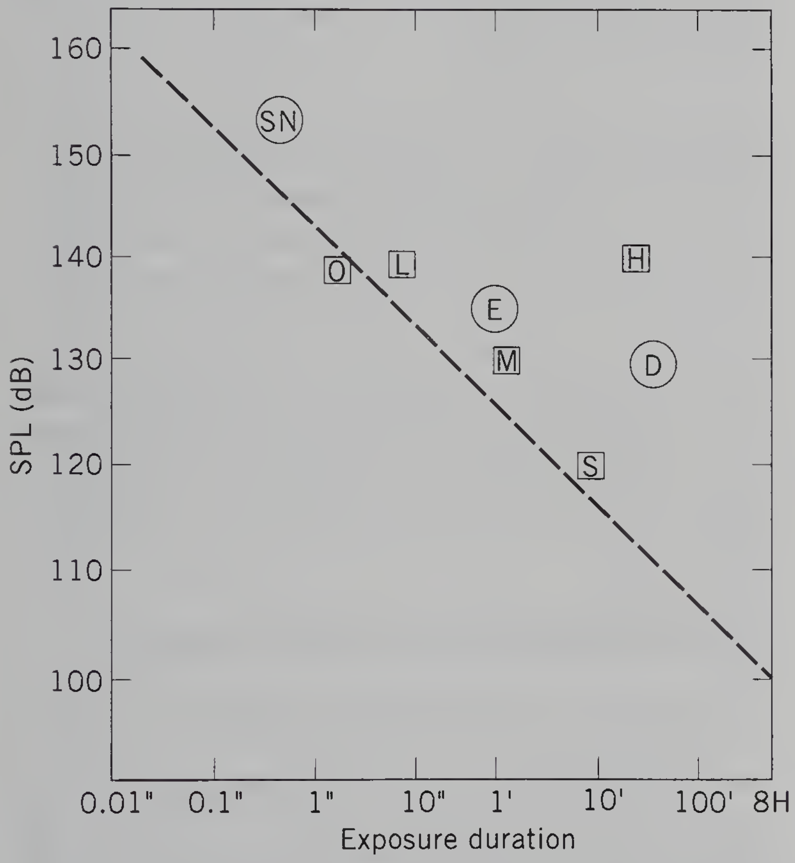


Fig. 3 Comparison of single noise exposures that have been shown to be without hazard to the average young healthy ear (circled symbols) and those that have apparently produced some permanent hearing loss in at least one person (symbols in squares). The dashed line, representing 8 h of exposure at 100 dB SPL or its energy equivalent, divides single exposures that are usually innocuous from those that are capable of causing permanent damage in the most highly susceptible individuals. See text for identification of data sources.

In Fig. 3, the spectrum of the various sounds is not considered. However, the difference between probably safe and possibly damaging points would be reduced if the ordinate were changed to "octave-band SPL," as the circled points would be reduced by a few decibels. In this case, the data are not in conflict with the generalization that acoustic trauma in humans is highly unlikely unless the single uninterrupted exposure exceeds 145 dB for 1 s or its energy equivalent, which is defined by the dashed line in Fig. 3.

Thus for single continuous steady exposures, the energy theory seems tenable: Exposures of equal energy give rise to approximately equal effects. Unfortunately, this equivalence breaks down for multiple exposures or a fluctuating sound level.

Gradual Growth of PTS Even more common and less understood than acoustic trauma is the gradual growth of hearing loss that accompanies repeated moderate exposures to sound, exposures that apparently only produce TTS. Cross-sectional studies of workers exposed daily to steady high-level industrial noise show that their average hearing is less sensitive than that of individuals who have quiet workplaces but whose age and exposure to noises outside the work situation are the same as for the noise-exposed workers. This difference in HTLs, which is by definition the INIPTS (industrial-noise-induced permanent threshold shift), grows rapidly with exposure time but reaches an asymptote in a few years. The terminal PTS increases linearly as level exceeds some base value (a level analogous to the EQ in the development of TTS) and is greater for the 3–6-kHz region than for lower frequencies.

Figure 4, from a synthesis by Passchier-Vermeer[44] of the literature involving continuous steady noises, shows the inferred INIPTS at the various audiometric frequencies as a function of workplace A-weighted sound level after 10 years of exposure. As a first approximation, it appears that (1) 80 dB(A) is the level that can be regarded as being innocuous; (2) 85 dB(A) will result

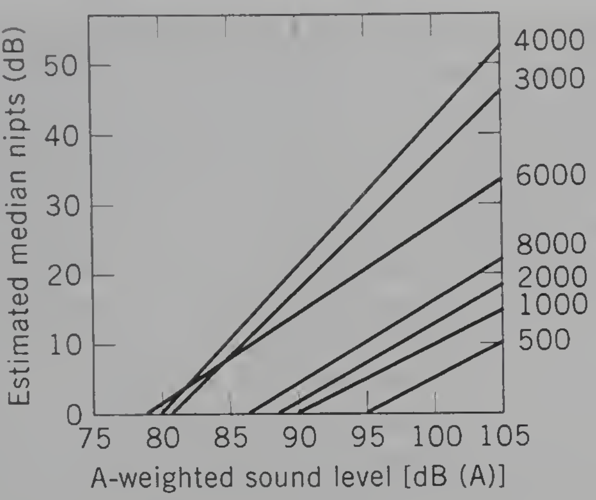


Fig. 4 Estimated average noise-induced permanent threshold shifts (ordinate) at various frequencies (parameter) after 10 years of exposure (8 h/day, 5 days/week) to steady A-weighted sound levels of 80–105 dB (abscissa). (After Ref. 44.)

in a loss of about 10 dB at the most noise-sensitive frequencies of 3, 4, and 6 kHz (10 dB is the smallest loss in the individual case that could be regarded as meaningful); (3) 90 dB(A) will generate a loss of 15–20 dB at these frequencies, although the traditional speech frequencies (0.5, 1, and 2 kHz) are still largely unaffected; only above 90 dB(A) does the noise adversely affect these lower frequencies; and (4) noises of 100 dB(A), which are still common, will produce severe losses at high frequencies and moderate losses at the low frequencies.

Thus Fig. 4 implies that 8 h exposure to a 90-dB(A) noise (plus, of course, whatever noise exposure is experienced outside the work situation) is capable of producing PTS when repeated daily for many weeks, even though the TTS at the end of a single workday will ordinarily recover completely. The value of TTS_2 of 17 dB that was estimated earlier to be the result of 8 h exposure to 90 dB(A) is just about equal to the 20 dB of PTS indicated in Fig. 4. This fact has lent support to the hypothesis that a daily exposure that produces a given TTS_2 in normal ears will eventually produce that same amount of PTS.[45] However, whether this agreement represents mere coincidence or proves a causal relation is still unclear.

There are two main hypotheses about how PTS grows with time in the individual. The more well-known theory posits that each day's exposure causes a change in the cochlea, but a change so small that it is undetectable by the rather gross and inherently variable audiometric procedures. This could be characterized as the *cumulative microtrauma* theory.[46] A 20-dB INIPTS displayed by a worker after 2 years of working in a 90-dB(A) noise (500 working days) could be regarded as the cumulative effect of 500 equal exposures, each of which produced a loss of gradually decreasing magnitude (e.g., the first exposure producing a loss of 0.30 dB, the second a loss of 0.28 dB, and so on, until perhaps the 500th exposure added only 0.001 dB).

An equally plausible scenario, however, rests on the fact that those 500 daily exposures were not identical. The exposure will vary from day to day not only during the 8-h workday but also during the intervening 16 h. It is possible, then, that damage will increase only when the total daily or weekly exposure is for some reason more severe than any preceding one. This might be termed the *progressive macrotrauma* theory.

Which of these two theories is more valid can be determined only by longitudinal studies in which the growth of PTS or some other indicator of permanent damage can be followed in individual experimental animals whose daily exposures can be held constant. At the moment, no clear decision has been reached; a study of guinea pigs indicated a gradual progressive growth of damage,[47] while chinchillas showed as much PTS after 1 week of exposure as after an additional 8 weeks.[48]

A conclusion regarding INIPTS that has been studied extensively enough that it can be regarded as a true principle is that INIPTS does not increase if noise exposure ceases. That is, NIPTS will not increase if a worker is removed from the noise.

Interrupted and Intermittent Exposures The damage produced by a schedule in which daily exposures of up to 8 h are separated (interrupted) by recovery periods of 16 h or more is less than that produced by an uninterrupted exposure. A further reduction of PTS in the chinchilla is effected if the daily exposure itself is intermittent rather than continuous,[49] just as in the case of TTS, although the reduction is not as great. The reduction in damage is greater, the shorter the noise bursts or the longer the intervening periods of relative quiet, although the former appears to be the more important factor.

3.3 Other Measures of Auditory Damage

As indicated earlier, physiological damage to the cochlea by high-intensity sound is not necessarily reflected in a measurable PTS. Evidence from both animal and human studies[50] implies that several hundred of the hair cells that have been presumed to be important in the process of hearing may be destroyed before a change in threshold is measurable. This clearly casts doubt on the adequacy of threshold sensitivity as the only valid index of damage. To date, however, efforts to find an indicator that is more sensitive to physiological damage have been unsuccessful, although many have been studied, including changes in the loudness, pitch, or maskability of suprathreshold tones, in difference limens for frequency and intensity, in temporal integration, and in strength of cochlear emissions. Although these indices are often abnormal when a threshold shift, either temporary or permanent, can be measured, they appear to return to normal if the threshold does.

Tinnitus A frequent aftereffect of intense stimulation is a "ringing in the ears." A tinnitus of this sort follows almost any impulse noise such as gunfire whose peak levels are above 150 dB; its duration ranges from a few seconds to several hours, depending on the severity of the exposure. Permanent tinnitus often but not universally accompanies acoustic trauma, and it may occur in an ear whose thresholds appear essentially normal.

3.4 Other Physiological Aftereffects

Cardiovascular System As mentioned earlier, there is only equivocal evidence that the changes in the cardiovascular system associated with high sound levels eventually lead to a significant enduring effect. While some epidemiological studies indicate a significant correlation between history of noise exposure or degree of hearing loss and hypertension or blood cholesterol content, such a relation has never been demonstrated when high-intensity sound is the only stressor involved. For example, although the incidence of cardiovascular disease was found to be higher than normal in steel mill workers,[51] their working environment included high temperature and noxious fumes as well as high sound levels, so the contribution to the coronary problems by the noise alone is indeterminate.

Miscellaneous Similarly, there is no convincing evidence that high-level sound alone can directly or even indirectly cause gastric ulcers, narrowing of the visual field, insomnia, abnormal somnolence, sexual impotence, birth defects, insanity, and death, inter alia, despite claims of antinoise alarmists whose allegations are amplified by the popular press. Of all of mankind's afflictions, only cancer and Alzheimer's disease seem never to have been attributed to noise exposure.

3.5 Susceptibility and Vulnerability

A given exposure, however it is measured, does not produce the same hearing loss in every ear. A particular ear may be more or less susceptible to damage—either to PTS or to TTS, which is assumed to be a precursor of PTS—because of relatively fixed structural and biochemical characteristics of the hearing mechanism and also more or less vulnerable to damage at a particular time because of temporary changes in one or more of those characteristics.

It is obvious that any factor that determines how much sound at a particular frequency reaches the inner ear will influence susceptibility, such as the dimensions of the pinna and the length of the auditory canal, the area and density of the eardrum, the mass of the ossicular chain, and the strength of the middle-ear muscles. Indeed, a conductive hearing loss renders the ear less susceptible to noise-induced hearing loss.[52] In the cochlea itself, one would expect characteristics such as the stiffness of the cochlear partition, the thickness and elasticity of the various membranes, and the adequacy of the blood supply, if these features could be measured, to be important in determining susceptibility to damage, either directly or, through their influence on recovery processes, indirectly. For example, there is some evidence that blue-eyed persons display higher thresholds than brown-eyed persons, which is explained if the melanin in the eye is highly correlated with the melanin in the cochlea and if the latter makes the cochlea more rigid.

Vulnerability depends on more labile characteristics of the auditory system. For example, an ear is less vulnerable to damage from sound when ear wax is blocking the ear canal or when a common cold has caused accumulation of fluid in the middle ear.[53] Vulnerability is higher in persons taking certain drugs that are themselves potentially ototoxic, such as kanamycin, neomycin, and streptomycin.

Although neither a particular exposure to noise nor a specific small dose of a drug may produce a measurable loss of hearing, their joint action may do so. Such synergistic action with noise has been demonstrated for only a few other substances or conditions, although dozens have been studied. Vulnerability to hearing damage seems to be affected by heat, cisplatin, aspirin, simultaneous whole-body vibration, and inhalation of carbon monoxide, although only the last named has been demonstrated to produce a true synergistic action.[54] Although attempts have been made to link susceptibility and vulnerability to general health, smoking habits, use of social drugs or stimulants, artificial food additives, industrial chemicals of various kinds, exercise, poor posture, and even mental attitude toward the source of sound, no reliable relation has yet emerged from these studies.[55]

A search for medications or procedures that will reduce TTS and PTS has been similarly unproductive. Although claims have been made that substances such as vitamin A, vitamin C, nicotinic acid, procaine, nylidrin, adenosine triphosphate, brain cortex gangliosides, gingko-biloba extract, amino-oxyacetic acid, and dextran have reduced PTS from acoustic trauma, subsequent studies failed to confirm this amelioration. The most promising substance at this time is carbogen (95% oxygen and 5% carbon dioxide), whose inhalation has been shown to reduce both TTS and PTS.[56,57]

Attempts to predict susceptibility by means of a single test in order to identify persons who will be at greatest risk have met with only limited success. It has always seemed reasonable to expect that the best predictor of PTS caused by extended exposure to a particular noise should be the TTS produced by a short exposure to that same noise, because both seem to depend on the same factors. However, studies in which both TTS and PTS were measured have generally shown correlations that, although statistically significant, are so low that they are of only minor practical importance, TTS accounting for only 15% of the variance in PTS among individuals.[58]

Although women on average have better hearing than men, even when working in the same noise, it is not clear whether this difference is due to some sex-linked structural difference or merely reflects differences in exposure to high-intensity sounds outside the workplace (socioacusis) and to other deleterious influences such as blows to the head and diseases (nosoacusis). In any event, the difference is not great enough to justify recommending that only women be employed in noisy occupations.

Age, experience, and hearing status all appear to be irrelevant as predictors of susceptibility. Despite many studies of these characteristics, it has not yet been demonstrated that the older person is more or less susceptible than the younger person, that those who work in noise gradually become more resistant to damage, or that workers with damaged hearing become more susceptible to further damage because of the injury.

REFERENCES

1. T. L. Wiley and R. S. Karlovich, "Acoustic-Reflex Response to Sustained Signals," *J. Sp. Hear. Res.*, Vol. 18, 1975, pp. 148–157.
2. W. D. Ward, "Studies on the Aural Reflex. III. Reflex Latency as Inferred from Reduction of Temporary Threshold Shift from Impulses," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 1132–1137.
3. G. R. Price, "Upper Limit to Stapes Displacement: Implications for Hearing Loss," *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 195–197.
4. J. J. Guinan, Jr. and W. T. Peake, "Middle-Ear Characteristics of Anesthetized Cats," *J. Acoust. Soc. Am.*, Vol. 41, 1967, pp. 1237–1261.
5. G. von Bekesy, "Zur Physik des Mittelohres und über das Hören bei fehlerhaften Trommelfell," *Akust. Z.*, Vol. 1, 1936, pp. 13–23.
6. R. C. Bilger and I. J. Hirsh, "Masking of Tones by Bands of Noise," *J. Acoust. Soc. Am.*, Vol. 28, 1956, pp. 623–630.
7. A. Finck, "Low-Frequency Pure Tone Masking," *J. Acoust. Soc. Am.*, Vol. 33, 1961, pp. 1140–1141.
8. J. C. Webster, "Noise and Communication," in D. M. Jones and A. J. Chapman (Eds.), *Noise and Society*, Wiley, London, 1984, pp. 185–220.
9. I. Pollack and J. M. Pickett, "Masking of Speech by Noise at High Sound Levels," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 127–130.
10. R. C. Bilger, "Remote Masking in the Absence of Intra-aural Muscles," *J. Acoust. Soc. Am.*, Vol. 39, 1966, pp. 103–108.
11. W. D. Ward, "Further Observations on Contralateral Remote Masking and Related Phenomena," *J. Acoust. Soc. Am.*, Vol. 42, 1967, pp. 593–600.
12. M. C. Liberman, "Structural Basis of Noise-Induced Threshold Shift," in *Proceedings of the Fifth International Congress on Noise as a Public Health Problem*, Vol. 4, Stockholm, Sweden, 1990, pp. 17–30.
13. C. Landis and W. A. Hunt, *The Startle Pattern*, Farrar and Rinehart, New York, 1939.
14. R. C. Davis, A. M. Buchwald, and R. W. Frankman, "Autonomic and Muscular Stimuli and Their Relation to Simple Stimuli," *Psych. Monogr.*, Vol. 69, No. 405, 1955.
15. R. Rylander and A. Dancer, "Startle Reactions to Simulated Sonic Booms: Influence of Habituation, Boom Level and Background Noise," *J. Sound Vib.*, Vol. 61, 1978, pp. 235–243.
16. G. Jansen, "Physiological Effects of Noise," in C. M. Harris (Ed.), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York, 1991, pp. 25.1–25.19.
17. M. Loeb, *Noise and Human Efficiency*, Wiley, New York, 1986, pp. 140–161.
18. S. R. Silverman, "Tolerance for Pure Tones and Speech in Normal and Defective Hearing," *Ann. Otol. Rhinol. Laryngol.*, Vol. 56, 1947, pp. 658–677.
19. R. A. Bentler and C. V. Pavlovic, "Comparison of Discomfort Levels Obtained with Pure Tones and Multitone Complexes," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 126–132.
20. H. W. Ades, A. Graybiel, S. Morrill, G. Tolhurst, and J. Niven, "Non-auditory Effects of High Intensity Sound Stimulation on Deaf Human Subjects," *J. Aviat. Med.*, Vol. 29, 1958, pp. 454–467.
21. E. D. D. Dickson and D. L. Chadwick, "Observations on Disturbances of Equilibrium and Other Symptoms Induced by Jet Engine Noise," *J. Laryngol. Otol.*, Vol. 65, 1951, pp. 154–165.
22. C. S. Harris, "Effects of Increasing Intensity Levels of Intermittent and Continuous 1000 Hz Tones on Asymmetrical Stimulation," *Perc. Motor Skills*, Vol. 35, 1972, pp. 395–405.
23. P. N. Borsky, "Review of Community Response to Noise," in *Proceedings of the Third International Congress on Noise as a Public Health Problem*, Freiburg, West Germany, ASHA Reports No. 10, 1980, pp. 453–474.
24. T. Schultz, "Synthesis of Social Surveys on Noise Annoyance," *J. Acoust. Soc. Am.*, Vol. 64, 1978, pp. 377–405.
25. W. D. Ward and D. Henderson, "The Dependence of Critical Level on Time," in *Proceedings of the Twelth International Congress on Acoustics*, Vol. 1, Toronto, Canada, July 24–31, 1986, N-4.3.

26. International Standardization (ISO), “Acoustics—Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment,” ISO 1990-01-15, ISO, Geneva, Switzerland, 1990.
27. W. D. Ward, “Adaptation and Fatigue,” in J. Jerger (Ed.), *Modern Developments in Audiology*, Academic, New York, 1973, pp. 301–344.
28. O. Bentzen, “Investigations on Short Tones,” Ph.D. Dissertation, Aarhus University, Denmark, 1953.
29. S. Kemp and R. N. George, “Masking of Tinnitus Induced by Sound,” *J. Speech Hear. Res.*, Vol. 35, 1992, pp. 1169–1179.
30. W. D. Ward, “Recovery from High Values of Temporary Threshold Shift,” *J. Acoust. Soc. Am.*, Vol. 32, 1960, pp. 497–500.
31. W. D. Ward, E. M. Cushing, and E. M. Burns, “Effective Quiet and Moderate TTS: Implications for Noise Exposure Standards,” *J. Acoust. Soc. Am.*, Vol. 59, 1976, pp. 160–165.
32. J. H. Mills, R. M. Gilbert, and W. Y. Adkins, “Temporary Threshold Shifts in Humans Exposed to Octave Bands of Noise for 16-24 Hours,” *J. Acoust. Soc. Am.*, Vol. 65, 1979, pp. 1238–1248.
33. W. D. Ward, A. Glorig, and D. L. Sklar, “Temporary Threshold Shift Produced by Intermittent Exposure to Noise,” *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 791–794.
34. H. McRobert and W. D. Ward, “Damage-Risk Criteria: The Trading Relation between Intensity and the Number of Nonreverberant Impulses,” *J. Acoust. Soc. Am.*, Vol. 53, 1973, pp. 1297–1300.
35. W. D. Ward, W. Selters, and A. Glorig, “Exploratory Studies on Temporary Threshold Shift from Impulses,” *J. Acoust. Soc. Am.*, Vol. 33, 1961, pp. 781–793.
36. W. D. Ward, “Temporary Threshold Shift and Damage-Risk Criteria for Intermittent Noise Exposures,” *J. Acoust. Soc. Am.*, Vol. 48, 1970, pp. 561–574.
37. A. Peyser, “Gesundheitswesen und Krankenfürsorge. Theoretische und experimentelle Grundlagen des persönlichen Schallschutzes,” *Deut. Med. Wochenschr.*, Vol. 56, 1930, pp. 150–151.
38. J. C. Saunders, Y. E. Cohen, and Y. M. Szymko, “The Structural and Functional Consequences of Acoustic Injury in the Cochlea and Peripheral Auditory System: A Five Year Update,” *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 136–146.
39. H. Davis, C. T. Morgan, J. E. Hawkins, Jr., R. Galambos, and F. W. Smith, “Temporary Deafness Following Exposure to Loud Tones and Noise,” *Acta Otolaryngol.*, Suppl. 88, 1950.
40. K. M. Eldred, W. J. Gannon, and H. Von Gierke, “Criteria for Short Time Exposure of Personnel to High Intensity Jet Aircraft Noise,” WADC Technical Note 55-355, Wright Air Development Center, U.S. Air Force, Wright-Patterson AFB, OH, September 1955.
41. H. C. Sommer and C. W. Nixon, “Primary Components of Simulated Air Bag Noise and Their Relative Effects on Human Hearing,” AMRL-TR-73-52, Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH, November 1973.
42. D. J. Orchik, D. R. Schraier, J. J. Shea, Jr., J. R. Emmett, W. H. Moreta, and J. J. Shea III., “Sensorineural Hearing Loss in Cordless Telephone Injury,” *Otolaryngol. Head Neck Surg.*, Vol. 96, 1987, pp. 30–33.
43. T. Lenarz and J. Gülzow, “Akustisches Innenohrtrauma bei Impedanzmessung. Akutes Schalltrauma?” *Laryngol. Rhinol.*, Vol. 62, 1983, pp. 58–61.
44. W. Passchier-Vermeer, “Hearing Loss Due to Exposure to Steady-State Broadband Noise,” *J. Acoust. Soc. Am.*, Vol. 56, 1974, pp. 1585–1593.
45. A. Glorig, W. D. Ward, and J. Nixon, “Damage Risk Criteria and Noise-Induced Hearing Loss,” *AMA Arch. Otolaryngol.*, Vol. 74, 1961, pp. 413–423.
46. D. W. Granendeel and R. Plomp, “Micro-noise Trauma?” *Arch. Otolaryngol.*, Vol. 71, 1960, pp. 656–663.
47. J. Herhold, “Über tierexperimentelle Untersuchungen zur Beurteilung der gehörschädigenden Wirkung des Lärms,” Ph.D. Dissertation, Technische University, Dresden, 1979.
48. Z. Zhou and W. D. Ward, “Is the Damaged Ear More Susceptible?” *J. Acoust. Soc. Am.*, Vol. 87, Suppl. 1, 1990, p. S102.
49. W. D. Ward, “The Role of Intermittance in PTS,” *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 164–169.
50. G. Bredberg, “Cellular Pattern and Nerve Supply of the Human Organ of Corti,” *Acta Otolaryngol.*, Suppl. 236, 1968.
51. G. Jansen, “Vegetative Lärmwirkungen bei Industriearbeitern,” *Lärmbekampfung*, Vol. 6, 1962, pp. 126–128.
52. R. Nilsson and E. Borg, “Noise-Induced Hearing Loss in Shipyard Workers with Unilateral Conductive Hearing Loss,” *Scand. Audiol.*, Vol. 12, 1983, pp. 135–140.
53. D. Y. Chung, “The Effect of Middle Ear Disorders on Noise-Induced Hearing Loss,” *J. Am. Audit. Soc.*, Vol. 4, 1978, pp. 77–80.
54. L. D. Fechter, J. S. Young, and L. Carlisle, “Potentiation of Noise Induced Threshold Shifts and Hair Cell Loss by Carbon Monoxide,” *Hearing Res.*, Vol. 34, 1988, pp. 39–48.
55. W. D. Ward, “Endogenous Factors Related to Susceptibility to Damage from Noise,” in *Occupational Medicine: State of the Art Reviews: Occupational Hearing Loss*, Hanley and Belfus, Philadelphia, 1995.
56. S. S. Joglekar, D. M. Lipscomb, and G. E. Shambaugh, Jr., “Effects of Oxygen Inhalation on Noise-Induced Threshold Shifts in Humans and Chinchillas,” *Arch. Otolaryngol.*, Vol. 103, 1977, pp. 574–578.
57. M. Hatch, M. Tsai, M. J. Larouere, A. L. Nuttall, and J. M. Miller, “The Effects of Carbogen, Carbon Dioxide, and Oxygen on Noise-Induced Hearing Loss,” *Hear. Res.*, Vol. 56, pp. 265–272.
58. W. Burns and D. W. Robinson, *Hearing and Noise in Industry*, Her Majesty’s Stationery Office, London, 1970, pp. 183–210.

120

CLINICAL AUDIOLOGY: AN OVERVIEW

LARRY E. HUMES

1 INTRODUCTION

Approximately 9% of the U.S. population is hearing impaired.[1] The prevalence of hearing impairment in this country, moreover, varies considerably with age. Among those 65 years of age and older, for example, approximately 30–40% have a significant hearing impairment.[1] Furthermore, the 65-and-over age group represents one of the most rapidly growing segments of the U.S. population. By the year 2000, it is estimated that one of every five individuals living in this country will be 65 years of age or older.[2] Given the rapid growth of the elderly in this country and the high prevalence of hearing impairment among this segment of the population, it is clear that the problems of the hearing impaired will demand more attention from those working in acoustics in the years ahead.

The problems of the hearing impaired have been the focus of a professional discipline within the broad realm of acoustics for the past 50 years. This field, known as audiology, was born during the 1940s as the result of national efforts to deal with the large number of hearing-impaired American soldiers returning home from World War II. The field of audiology has undergone rapid growth and diversification in the intervening decades, but the primary focus of this field remains the rehabilitation of those with hearing impairment. Today, there are approximately 11,000 audiologists in this country working in settings that include public schools, hospitals, speech and hearing clinics, industry, and independent private practices.

Although clinical audiology is an increasingly diverse profession, there is a common core of information and skills shared by all audiologists. It is this common core that is reviewed in the remaining pages of this chapter. The remainder of this chapter has been divided into two sections: (1) diagnostic audiology and (2) rehabilitative audiology.

2 DIAGNOSTIC AUDIOLOGY

One of the first steps in the rehabilitation of the hearing impaired is the establishment of an accurate diagnosis. The diagnosis is typically established with the assistance of a physician. The audiologist conducts a wide array of auditory tests to assist the physician in determining the extent of hearing difficulty and its underlying cause or etiology. A critical piece of information used in establishing the etiology of hearing impairment is determination of the location of the pathology within the auditory system. This is generally referred to as establishing the *site of lesion*. Much information about the extent of the impairment and its cause can be obtained from the tests included in the basic hearing evaluation, which is comprised of three types of tests: (1) pure-tone audiometry, (2) immittance measurements, and (3) speech audiometry.

2.1 Pure-Tone Audiometry

In pure-tone audiometry, sinusoidal signals presented to the listener's ears are adjusted in sound pressure level (SPL) to determine the level that can just be detected 50% of the time. These behavioral measurements are made for each ear at octave intervals from 250 through 8000 Hz using standardized transducers to deliver the stimuli to each ear. The most common types of trans-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

ducer used clinically today are supra-aural earphones, insert earphones, and bone vibrators. When the sinusoidal stimuli are delivered with either of the first two electroacoustic transducers an acoustic signal is delivered to the outer ear of the listener and the term *air conduction* testing is used. For such signals to be perceived normally, the entire peripheral auditory system, from the outer ear through the auditory nerve, must be functioning normally.

The bone vibrator, on the other hand, is an electromechanical transducer that is usually coupled with a headband to the bony skull at the forehead or to the mastoid process of the temporal bone (the bony protuberance just behind the outer ear). At either location, the bone vibrator stimulates the bony skull surrounding the inner ear and (in a normal ear) gives rise to an auditory sensation identical to that of air-conducted stimuli. By stimulating the inner ear directly, the bone conduction measurements have bypassed the more peripheral outer and middle ears. Differences in the sound levels needed to detect a sinusoid via air and bone conduction can offer some insight into the location of pathology underlying a particular hearing loss.

The just-detectable sound levels, or hearing thresholds, measured at each frequency by air and bone conduction are plotted on a standardized graph referred to as an audiogram. The top panel of Fig. 1 displays an audiogram from both ears of a young adult with normal hearing, whereas the bottom panel displays the audiogram from both ears of an "average" 65-year-old male. Notice that the normal-hearing young adult has an audiogram that is near 0 dB across the entire range of frequencies. The decibel scale used on an audiogram is referred to as decibels hearing level (HL). The just-detectable SPLs at each frequency represent the reference for this scale. (That is, dB HL = dB SPL_m − dB SPL_{ref}, or 20 log P_m − 20 log P_{ref}, where the subscripts m and ref refer to the patient's measured values and standardized reference values, respectively.) Table 1 displays the reference threshold SPLs at each frequency that correspond to 0 dB HL on the audiogram. It is important to emphasize that the values in this table represent transducer-specific average threshold SPLs derived from several studies and hundreds of "otologically normal" adults (free of ear disease). Thus, this normal-hearing young adult with thresholds of 0 dB HL at all frequencies requires 45 dB SPL at 250 Hz and 7 dB SPL at 1000 Hz to detect these sinusoidal signals.

Notice that increasing hearing loss in decibels hearing level is plotted on the audiogram by moving down on the graph. Thus, contrary to what one might expect, more intense sounds are plotted *lower* on the graph than less intense sounds. (If considered in terms of hearing *loss*, however, the y-axis may not appear to be so counterintuitive.) The audiogram in the lower panel of Fig. 1, then, indicates that this 65-year-old man has more hearing loss in the high frequencies than in the low frequencies.

TABLE 1 Standard Reference Equivalent Threshold SPLs for a TDH-39 Earphone Measured on the NBS-9A Coupler

Frequency (Hz)	dB SPL
125	45.0
250	25.5
500	11.5
1000	7.0
2000	9.0
4000	9.5
8000	13.0

Source: From Ref. 3.

The audiograms in Fig. 1, moreover, serve to illustrate several other features of audiogram notation. First, a standardized set of symbols is used to graph the results. Air conduction (AC) thresholds from the right ear (RE), for example, are represented by open circles, whereas "X" symbols are used for the corresponding results from the left ear (LE). The left- and right-pointing arrowheads shown in Fig. 1, moreover, represent the results from bone conduction (BC) testing obtained from the right and left ears, respectively. (For reasons not necessary to review here, an audiologist would actually seldom obtain BC thresholds from each ear in the cases illustrated in this figure. Results from the nontest ear can often be inferred from those obtained from the test ear via BC.)

The audiogram obtained from pure-tone audiometry provides several pieces of information at a glance. First, the extent and severity of the hearing impairment is depicted. That is, the amount of hearing loss and the frequencies affected are both readily apparent from the audiogram. Second, comparisons across ears can be made readily. In both audiograms shown in Fig. 1, the hearing thresholds were essentially equivalent in both ears. This is known as a *bilaterally symmetrical* hearing loss. (Measurement error is 5 dB in clinical pure-tone audiometry so that between-ear differences must exceed 5 dB to be considered real differences.) Third, any difference in hearing threshold measured by AC and by BC will be readily apparent. This difference, referred to as the *air–bone gap*, is of diagnostic significance. Bone conduction thresholds are never significantly worse than AC thresholds, but they can be significantly better. Better hearing by BC implies a lower (or better) hearing threshold for the inner ear compared to that of the entire auditory periphery (outer, middle, and inner ears). If a higher hearing threshold is measured for air-conducted sound than for bone-conducted sound, then the extra hearing

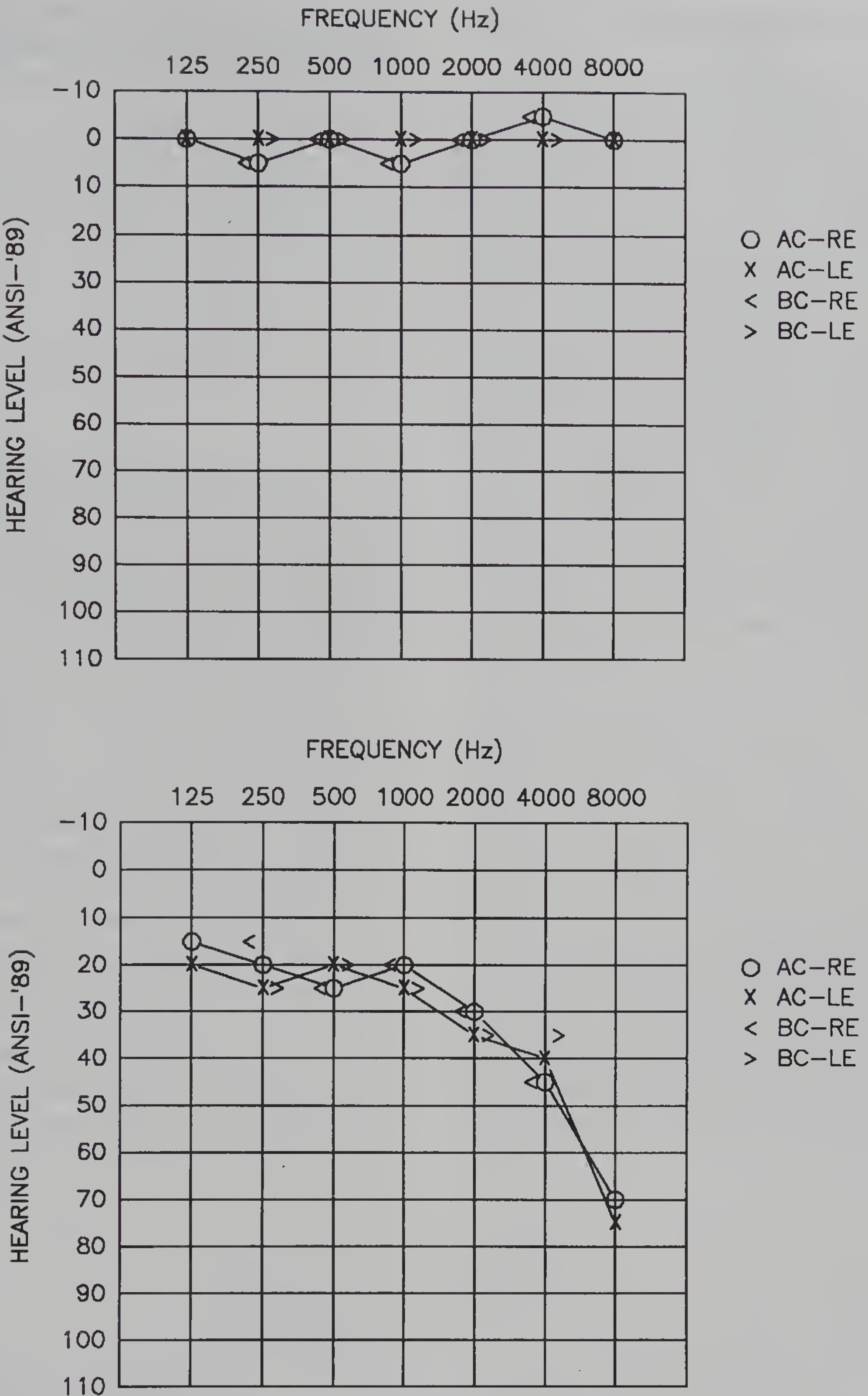


Fig. 1 Audiogram from a normal-hearing young adult (top) and a hearing-impaired 65-year-old male with the high-frequency sensorineural hearing loss (bottom) (AC, air conduction; BC, bone conduction; RE, right ear; LE, left ear).

loss associated with the air-conducted sound must be due to a disorder of the outer or middle ear. Air–bone gaps of 10 dB or more are considered significant clinically and indicate the presence of pathology in the outer or middle ear. Because the pathology usually attenuates the sound being "conducted" through the outer and middle ears to the inner ear, this type of hearing loss is referred to as *conductive* hearing loss. Examples of common causes of conductive hearing loss include excessive cerumen ("wax") in the outer ear, fluid in the middle ear associated with "ear infections" or "earaches" (otitis media), and ossification of the bones in the middle ear (otosclerosis). The hearing loss measured for AC signals in conductive pathology is usually flat across frequency or slightly worse in the low frequencies than in the high frequencies.

When hearing loss is present but the thresholds measured by air and bone conduction are equivalent (as in the lower panel of Fig. 1), the hearing loss is referred to as *sensorineural* hearing loss. Since the hearing threshold of the inner ear alone (bone conduction) is the same as that of the entire auditory periphery (air conduction), the presence of a sensorineural hearing loss implies a disorder of the sensory or neural portions of the inner ear.

The results of pure-tone audiometry could also reveal the presence of a hearing loss that was both conductive and sensorineural. This is referred to as a *mixed* hearing loss. If the 65-year-old male whose audiogram is shown in the bottom panel of Fig. 1 was retested after wax had accumulated in the outer ear, then the AC thresholds in the low and midfrequencies would be higher while the BC thresholds would remain the same. The appearance of this air–bone gap would indicate the presence of a conductive hearing loss superimposed on the existing sensorineural hearing loss.

In the top panel of Fig. 1, the normal-hearing young adult had hearing thresholds near 0 dB HL at all frequencies. There is actually a range of normal hearing from −10 to 25 dB HL. Although there is some debate about the upper limit of "normal" hearing, an individual with AC hearing thresholds less than or equal to 25 dB HL at all frequencies would be considered normal. Table 2 describes a classification system commonly used in audiology to categorize different degrees of hearing loss. According to this scheme, for example, the audiogram from the elderly individual in the bottom panel of Fig. 1 would represent a bilaterally symmetrical, moderately severe, high-frequency sensorineural hearing loss.

TABLE 2 Classification of Severity of Hearing Impairment Based on Air Conduction Pure-Tone Thresholds

Thresholds (dB HL)	Classification
−10–25	Normal
26–40	Mild
41–55	Moderate
56–70	Moderately severe
71–90	Severe
> 90	Profound

2.2 Immittance Measurements

Immittance measurements represent another major component of the basic hearing evaluation. Generally, these measurements are used to help establish the site of lesion within the auditory system and are not used to determine the extent or severity of hearing loss.

"Immittance" is actually a nonsensical term without a physical foundation that was adopted in an effort to create a uniform terminology for these measurements. The clinical instruments used to measure immittance actually measure some aspect of either the impedance of the middle ear or its reciprocal, admittance.

Immittance measurements are made with the assistance of a special transducer assembly that fits into the ear canal. Figure 2 illustrates a schematic cross-sectional diagram of the assembly sealed in the ear canal. Essentially, there are three tubes that terminate in the assembly. One tube is connected to a receiver used to deliver a sinusoidal signal into the sealed ear canal. A second tube is connected to a microphone used to measure the level

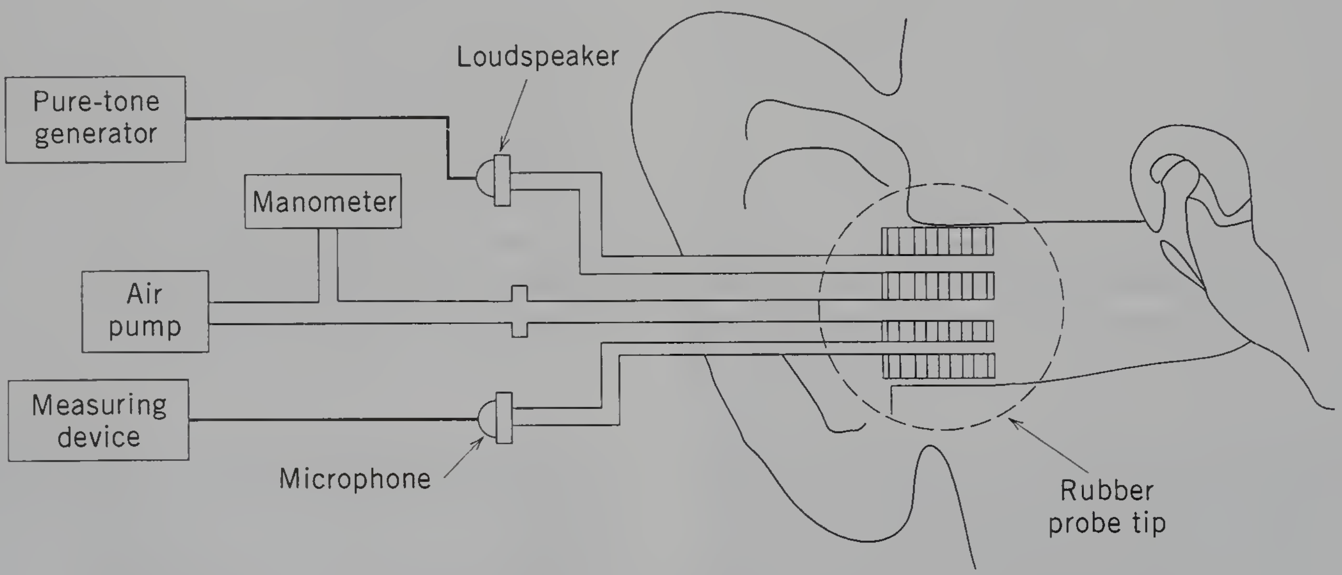


Fig. 2 Schematic cross section of the probe assembly of an immittance measurement system inserted in the ear canal.

of the sound in the sealed ear canal. Finally, a third tube is connected to a manometer and is used to vary the air pressure inside the sealed ear canal in a controlled fashion.

Several measurements are possible with this assembly, but the general principle behind them all is that as the impedance of the middle ear increases, less of the sound introduced into the ear canal by the receiver will be transmitted through the middle ear and more will be reflected back into the sealed ear canal cavity. The increased sound level in the sealed ear canal will be picked up by the microphone and displayed in one of several ways (needle deflection on a meter, pen displacement on a graphic level recorder, digital display, etc.).

One of the tests comprising the immittance test battery is typanometry. In tympanometry, the manometer is used to systematically vary the air pressure in the sealed ear canal from approximately −400 to +200 daPa. At the extreme negative and positive air pressures, the stiffness of a normal middle ear is increased and more sound is reflected back into the sealed cavity. This results in an increased sound level in the ear canal measured by the microphone. As the air pressure in the ear canal approaches that on the other side of the eardrum in the air-filled middle ear, the middle ear becomes more compliant and less sound is reflected back into the ear canal. This results in a decreased output from the microphone. When the output of the microphone is recorded graphically, the result is referred to as a tympanogram.

Figure 3 displays three tympanograms. The solid line represents a normal tympanogram. There are several features of the tympanogram used to define its normality, including the peak amplitude, the ear canal air pressure at which the peak occurs, and the general shape of the tympanogram. One common tympanogram abnormality is illustrated by the dashed tympanogram. Note that this tympanogram is nearly identical to that of the normal middle ear in amplitude and shape but is shifted horizontally so that the peak occurs at an air pressure of −200 daPa. This is referred to as negative middle-ear pressure and, if chronic, can be a precursor to the development of otitis media (ear infections or earaches). A short-lived, acute change in middle-ear pressure is frequently experienced by rapid changes in elevation, as in air travel. Finally, the dotted tympanogram in Fig. 3 illustrates a flat tympanogram associated with a fluid-filled middle ear, as in otitis media. The fluid has increased the stiffness of the middle ear and results in greater sound being reflected back into the sealed ear canal at all air pressures. It should be noted that because the vast majority of middle-ear disorders affect the stiffness of the middle ear, the sinusoid introduced into the sealed ear canal through the receiver tube is usually low in frequency (220 and 660 Hz are the most commonly used frequencies).

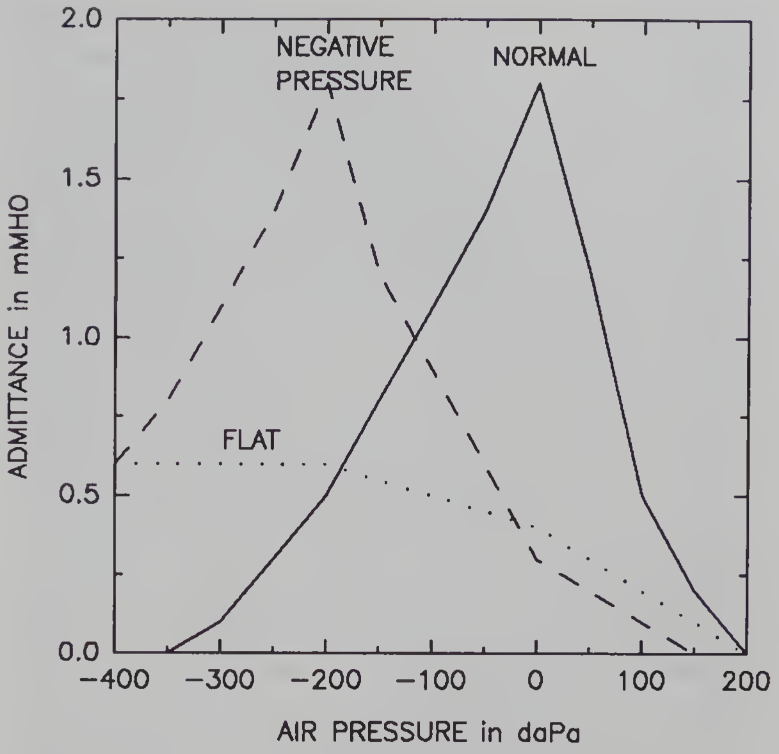


Fig. 3 Three tympanograms measured with an immittance measurement system. Tympanograms illustrated are normal (solid line), negative pressure (dashed line), and flat (dotted line).

Tympanograms are used to assess the integrity of the middle ear. Another component of the immittance test battery, however, is sensitive to the presence of pathology in the middle ear, inner ear, auditory nerve, and brainstem. This test, referred to as acoustic-reflex measurement, involves the reflexive contraction of the middle-ear muscles in response to high-intensity sound. The two aspects of acoustic-reflex measurements that are of primary clinical importance are reflex threshold and reflex decay. In the measurement of reflex threshold, the intensity of a stimulating sound is adjusted to determine the level that will just elicit a muscle contraction. The transducer assembly used with tympanometry is also used to monitor the middle ear for the occurrence of a muscle contraction. When the muscles contract, the stiffness of the middle ear is increased slightly, which increases the ear canal SPL measured by the microphone in the assembly. The acoustic stimulus delivered to the ear to elicit the reflex either can be presented to the same ear with the transducer assembly through a fourth tube or it can be delivered to the opposite ear with an earphone. These two different stimulus presentation modes are referred to as ipsilateral and contralateral presentations, respectively. Contralateral presentation is possible because the middle-ear reflex is a consensual phenomenon. That is, for a normal auditory system, stimulation of one ear with high-level sound results in middle-ear muscle contractions on both sides.

Complete details regarding the diagnostic significance of acoustic reflex thresholds are beyond the scope of this overview. Essentially, reflex thresholds are measured for pure tones at octave intervals from 500–4000 Hz and are judged as being within the normal limits (<105 dB HL), elevated (>100 dB HL), or absent. Elevated or absent acoustic reflexes can be observed in individuals with normal pure-tone hearing (likely indicating an auditory nerve or brainstem lesion) or in individuals with conductive or sensorineural hearing loss. Thus, elevated or absent acoustic reflex thresholds, in and of themselves, do not pinpoint the site of lesion. When interpreted in conjunction with the tympanograms and the pure-tone audiogram, however, acoustic reflex thresholds can offer considerable insight into the location of the pathology within the auditory system.

If an acoustic reflex threshold can be obtained for a 500- or 1000-Hz stimulus, then measures of reflex decay are also usually obtained at one or both of these frequencies. Using a suprathreshold sound level, the magnitude of immittance change associated with the reflex is monitored for a 10-s presentation of the reflex-activating sound. The magnitude of the decay or adaptation of the reflex is then quantified and classified as abnormal (too rapid decay) or normal. Abnormal reflex decay is usually an indication of a problem in the auditory nerve or brainstem, such as an acoustic neuroma (tumor on the auditory nerve).

2.3 Speech Audiometry

The third primary component of the basic hearing evaluation is speech audiometry. There are several variations of speech audiometry, but the two measurements common to this area of testing in the United States are speech recognition threshold (SRT) and word recognition score (WRS). Briefly, measurement of the SRT involves the presentation of standardized lists of two-syllable words ("spondees") at various intensities to determine the sound level corresponding to 50% correct recognition of these words. Word recognition performance, on the other hand, typically makes use of standardized lists of one-syllable consonant–vowel–consonant words (CVC monosyllables) presented at a fixed level. The level used is typically 40 dB above the SRT. Whereas the SRT determines the softest level at which speech can be recognized, the WRS estimates how well speech can be understood at a comfortable level.

Word recognition performance is often also assessed in a background of competition. The competition is frequently white noise, speech-shaped noise, or a multitalker babble. The speech-to-competition ratio may vary from 0 to +10 dB, depending on the application and the particular set of speech materials. In addition, speech materials may be presented in an open (fill-in-the-blank) or closed (multiple-choice) response format. Finally, speech materials other than monosyllables may be used. Standardized lists are available for nonsense syllables as well as for meaningful sentences. Although standardized test materials exist, the presentation level of the materials, as well as the type and level of background competition, have not been standardized.

2.4 Basic Hearing Evaluation: A Synopsis

It is always desirable to obtain as much information from a given hearing-impaired individual as possible. In the basic hearing evaluation, results are needed from pure-tone audiometry, immittance measurements, and speech audiometry. This information, along with a detailed case history, can go a long way toward establishing an accurate diagnosis. It is not always possible, however, to obtain measures from all of the tests in this basic hearing evaluation battery. For example, it is seldom possible to obtain an accurate and complete pure-tone audiogram from very young children (less than 2 years of age), although alternative procedures are available to the audiologist to permit the assessment of hearing in children as young as 3–6 months of age. Nonetheless, for the vast majority of cases seen by the audiologist, results from all three components of the basic hearing evaluation are readily obtained.

The results of individual components of the basic hearing evaluation are frequently mutually reinforcing and point to the same site of lesion. Consider the case illustrated in Fig. 4. The top panel of this figure shows the AC and BC thresholds from the right ear of an individual with chronic otitis media. This disorder is very common in children and is often characterized by the accumulation of fluid in the middle ear. The left ear (not shown here for clarity) is assumed in this example to be normal in all respects. Note that inner ear (BC) hearing in the right ear is normal whereas AC thresholds indicate the presence of a mild, flat hearing loss. These large air–borne gaps are consistent with the presence of conductive pathology in this ear.

The immittance measurements summarized in the lower panel of this figure are also consistent with the presence of conductive pathology in the right ear. Note that the tympanogram from this ear is flat and that contralaterally activated acoustic reflexes are absent bilaterally. When the high-intensity reflex-activating sound is presented to the normal left ear, the extreme stiffness of the fluid-filled right ear prevents the detection of the small change in stiffness associated with reflex

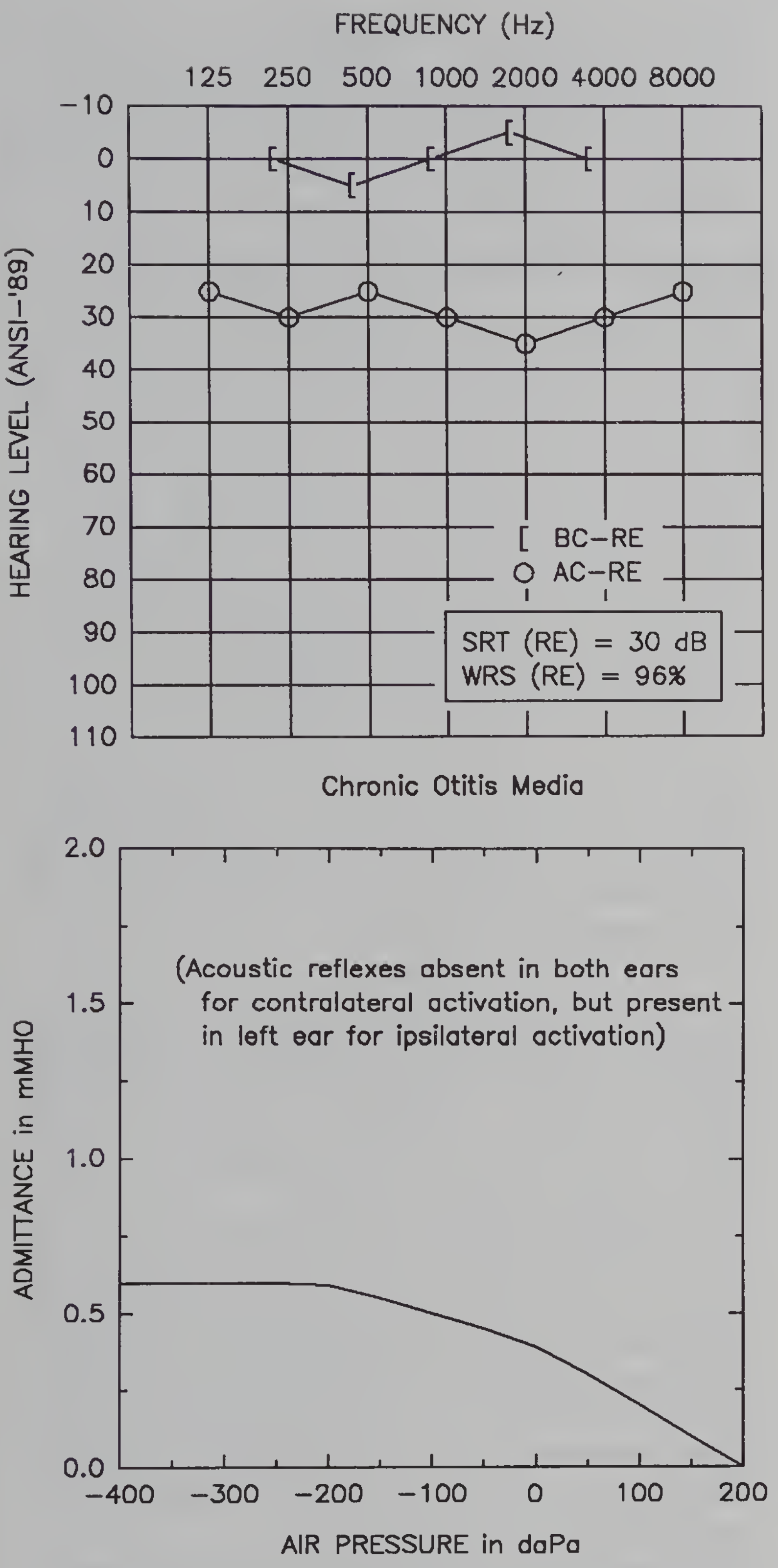


Fig. 4 Results of the basic hearing evaluation for a child with chronic otitis media in the right ear and normal hearing in the left (not shown). Results of pure-tone and speech audiometry are shown in the top panel and immittance measurements are shown in the bottom panel.

contraction. When the activating stimulus is presented to the right ear, however, the attenuation of the activating sound by the conductive hearing loss in the right ear results in a stimulus level that is below the reflex threshold and an acoustic reflex does not occur, even though the probe assembly is in the normal left ear. Ipsilateral presentation of the activating sound to the left ear does reveal the presence of a normal acoustic-reflex threshold, as expected for this healthy ear.

Finally, the results from speech audiometry are also consistent with the presence of conductive pathology in the right ear. The SRT, measured with headphones, agrees with the average hearing loss measured for pure tones delivered the same way (AC thresholds). In addition, as is true with most cases of conductive pathology, excellent WRSs are observed once the level of the speech is increased to compensate for the attenuation of the hearing loss. This is often *not* the case for many individuals with a sensorineural hearing loss.

The results of the basic hearing evaluation in this example all point to the same conclusion: the presence of conductive middle-ear pathology in the right ear. In many cases, however, the results from the basic hearing evaluation itself do not pinpoint the site of lesion. Advanced special tests, usually consisting of various measures of speech audiometry or measures of the electrical activity of the auditory portions of the brainstem and cortex evoked by acoustic stimuli, are frequently needed to determine the site of lesion.[4,5] Description of these advanced tests and their interpretation, however, is beyond the scope of this brief overview.

3 REHABILITATIVE AUDIOLOGY

As mentioned, the first step in the rehabilitation process is the determination of the extent and severity of hearing impairment and the site of lesion. In the vast majority of conductive hearing impairments, the underlying pathology can often be eliminated and hearing restored through medical treatment consisting of prescribed medication, surgery, or both. If the site of lesion is retrocochlear (pathology of the auditory nerve or auditory pathways in the brainstem or cortex), medical intervention (frequently surgery) is also necessitated. For those individuals with chronic untreatable middle-ear pathology or sensorineural hearing loss associated with a cochlear site of lesion, on the other hand, aural rehabilitation is the primary course of action.

A comprehensive review of aural rehabilitation is also beyond the scope of this overview of clinical audiology. Much of the early portion of the aural rehabilitation process, however, is centered on the selection and evaluation of a prosthetic device. The types of devices available, their function, and their advantages and limitations are described in the pages to follow. The types of devices reviewed, in the order of their presentation, are (1) conventional electroacoustic hearing aids, (2) assistive listening devices, (3) vibrotactile devices, and (4) cochlear implants. It should be emphasized, however, that the aural rehabilitation process *begins*, not *ends*, with the selection and evaluation of a prosthetic device. Following device selection, the hearing-impaired individual needs to participate in a variety of rehabilitation activities, including auditory training, training in effective communication strategies, and training in lipreading. Finally, the focus in this section is on the hearing-impaired adult. The rehabilitation process is considerably different for children, especially if the hearing loss is severe and occurs prior to the normal development of speech and language.

3.1 Conventional Electroacoustic Hearing Aids

The conventional electroacoustic hearing aid is essentially a miniature, head-worn, public address system, complete with microphone, amplifier, and loudspeaker (receiver). One of the primary differences between the typical hearing aid and a public address system is that the latter system places the microphone close to the desired sound source. This is not the case with the conventional hearing aid and, as will be demonstrated, remains one of the primary limitations of this system.

There have been significant advances in the conventional hearing aid over the past decade. The size of these instruments has been reduced from devices worn in the shirt pocket or on top of the ear to units that can be worn entirely inside the ear canal of the listener. At the same time, fidelity and electroacoustic flexibility have been enhanced. Many hearing aids available today offer directional microphones and various signal-processing options designed to reduce background noise. In addition, several contemporary conventional hearing aids are programmable, storing four to seven sets of electroacoustic characteristics in memory. Remote controls with infrared coupling to the hearing aid represent another recent advance that is of considerable benefit to elderly wearers lacking the manual dexterity to adjust the tiny controls on a head-worn hearing aid.

Most conventional hearing aids are selected for a particular hearing-impaired person on the basis of the extent and severity of the hearing loss. The primary problem experienced by most individuals with sensorineural hearing loss associated with a cochlear site of lesion is that speech can be heard, but not understood. The audiograms in Fig. 5 help to explain the reasons for this common complaint and how a well-fitted hearing aid can compensate for it. The "X" symbols on this audiogram represent the hearing thresholds from the left ear of a normal-hearing young adult. Superimposed on this audiogram is the approximate range of speech intensities at each frequency for speech at a conversational level (65 dB SPL at 1 m). This range, which encompasses the 10th to 90th percentiles of the amplitude distributions at each frequency, is represented by the dashed lines. Also shown are the approximate locations of various speech sounds plotted on these same coordinates. Note that all of the speech sounds are well above threshold for this hypothetical normal-hearing young adult.

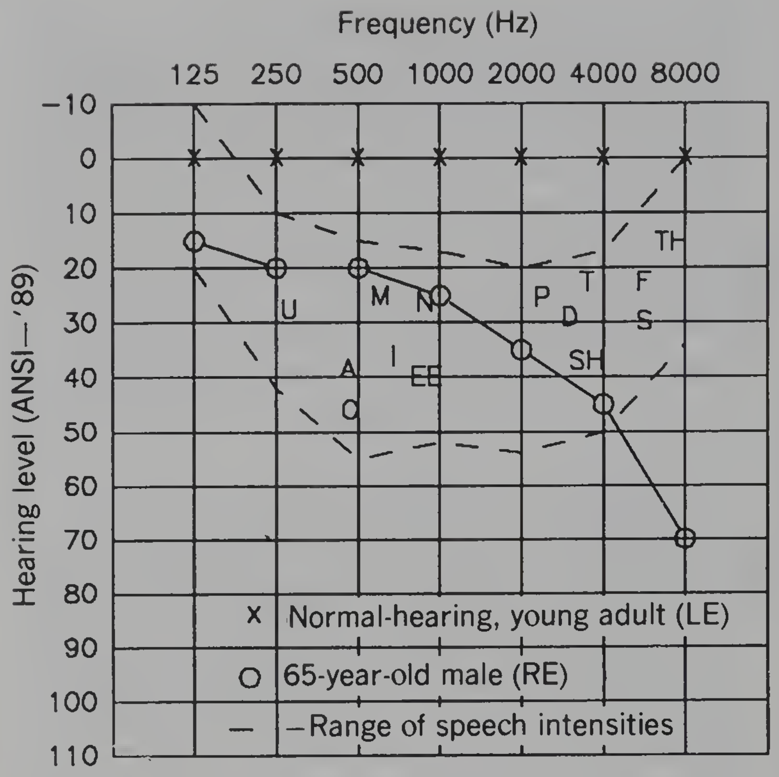


Fig. 5 Audiogram with the range of speech intensities (dashed lines) and various speech sounds superimposed on it. The crosses (×) represent hearing thresholds for a normal-hearing young adult and the circles represent hearing thresholds of a typical 65-year-old male.

Figure 5 also illustrates the audiogram from the right ear of an average 65-year-old male (circles). Note that the presence of a high-frequency sensorineural hearing loss renders many of the high-frequency, low-intensity consonants inaudible for this individual. Words such as *wife*, *white*, *wize*, *wipe*, and *wide* are essentially all heard by this listener as *why*. Thus, the listener *hears* speech when each word is spoken but cannot understand what has been said.

Consider now the effect of a well-fitted hearing aid on the audibility of speech. By selectively amplifying speech in the frequency region with hearing loss, the high-frequency consonants that were previously inaudible can now be heard comfortably by the listener. In the example shown in Fig. 5, for instance, little or no gain would be required below 1000 Hz, with increasing amounts of gain needed as frequency increased above 1000 Hz. Under these circumstances, a well-fitted conventional hearing aid typically results in large improvements in speech understanding.

There are three primary limitations, however, of a well-fitted conventional hearing aid that restrict the improvements that can be achieved. First, communication seldom takes place in quiet. Frequently there is noise or other conversation in the background. The conventional hearing aid amplifies the speech and the noise equally and does nothing to improve the speech-to-noise ratio. In fact, because of bandwidth limitations, distortion, and internal noise, the conventional hearing aid may often effectively *decrease* the speech-to-noise ratio. The incorporation of directional microphones and signal-processing circuits in conventional hearing aids in recent years represent two efforts to improve the speech-to-noise ratio.

The second limitation of conventional hearing aids has to do with the acoustic gain that can be achieved by the instrument without developing feedback. Acoustic feedback is a serious limitation for many hearing aids because of the close proximity of the microphone to the receiver, especially for the tiny instruments that fit entirely within the ear canal. Several feedback reduction circuits are currently being investigated as a means of alleviating this problem.

The third primary limitation has to do with the severity of hearing loss and the amount of amplification provided by the hearing aid. Although listeners with sensorineural hearing impairment cannot hear low-level sounds as well as normal-hearing listeners, high-intensity sounds are usually equally loud for both types of listener. Consequently, the amount of amplification required to make all of the low-intensity speech sounds audible cannot always be achieved without overamplifying other sounds of moderate or high intensity. The output of a hearing aid must be limited by peak clipping or amplitude compression to avoid presentation of sound levels that are uncomfortably loud or even potentially harmful to the hearing aid wearer. Amplitude compression continues to be investigated as an alternative to linear amplification with peak clipping. Compression schemes that have been investigated include variations in the number of channels, attack–release times, compression thresholds, compression ratios, and whether the system makes use of input- or output-dependent compression.[6,7]

The electroacoustic characteristics of a conventional hearing aid, such as its frequency–gain characteristics and its saturation SPL, are typically prescribed for a particular hearing-impaired listener using one of several available, theoretically based, prescriptive methods. The prescriptions are generated for a particular hearing-impaired person from knowledge of the individual's threshold and uncomfortable loudness level at several frequencies. The accuracy of the fit of the custom-ordered hearing aid to the prescription is then confirmed in the listener's ear canal using a tiny probe-tube microphone. Gain and maximum output characteristics are measured at several frequencies in a matter of a few minutes, and the measured characteristics are then compared to those prescribed for the individual and fine tuned to obtain a reasonable match.

3.2 Assistive Listening Devices

As mentioned, one of the primary limitations of the conventional hearing aid is the poor speech-to-noise ratio present at the microphone. To compensate for the inaudibility of some speech sounds, even with amplification, hearing-impaired listeners frequently require a better than normal speech-to-noise ratio. One simple way in which this can be accomplished is to move the microphone closer to the desired sound source, and many devices have been designed in recent years to do just that.[8] These devices, referred to as assistive listening devices, represent alternatives or supplements to conventional hearing aids. They must be considered as alternatives to hearing aids for those impaired listeners with mild hearing loss requiring only limited assistance of a specific nature or as supplements to hearing aids for severely hearing-impaired listeners requiring additional assistance beyond that available from their hearing aid.

As noted, one class of assistive listening device is designed to improve the speech-to-noise ratio by moving the microphone closer to the desired sound source. The desired sound source might be a talker, television, stereo, and so on. A detached microphone is placed in close proximity to the sound source, and its electrical output is sent to a body-worn amplifier and receiver. The output of the microphone can be sent to the amplifier via hardwire, FM transmission, or infrared transmission. The amplifier output can also be delivered to the ears of the listener in a variety of ways, including lightweight stereo headsets, stethoscope-type earphones, insert earphones, or hearing aids.

Other classes of assistive listening devices include special telephone listening aids that can be used with or without a hearing aid to improve telephone communication. Special visual or vibrotactile alerting devices are also available for profoundly impaired listeners. These devices are designed to flash lights that can be seen or produce vibration that can be felt by the hearing-impaired individual when specific acoustic events take place, such as the ring of a doorbell, an alarm clock, or a telephone. This latter type of device is clearly a supplement to the conventional hearing aid for those profoundly impaired individuals not receiving enough benefit from conventional amplification.

3.3 Vibrotactile Devices

For many individuals with profound sensorineural hearing loss in both ears, alternatives to conventional amplification have been made available that may generally be referred to as sensory substitution devices. That is, rather than delivering high-intensity sound to a damaged ear through a hearing aid or an assistive listening device, the sound is converted to an alternate form of energy and presented to the individual. For vibrotactile devices, the sound is converted to an electrical signal and delivered to the skin as a pattern of mechanical vibration by an array of vibrating mechanical contacts. Frequency can be encoded by both the rate of contact vibration and the location of the vibrating contact within the array. Amplitude may also be encoded by these devices in various ways, including the amplitude and rate of vibration. For most vibrotactile devices, the array of vibrating contacts is either placed on the forearm or incorporated into a belt worn around the midsection. An excellent overview of the relative advantages and disadvantages of conventional amplification, cochlear implants, and vibrotactile devices was provided recently by a working group of the Committee on Hearing, Bioacoustics and Biomechanics (CHABA) of the National Research Council.[9]

3.4 Cochlear Implants

Cochlear implants transduce the acoustical signal to an electrical one, process it, and deliver the electrical signal directly to the nerve fibers in the inner ear. Although a variety of implant types exist and each differs in some way from the others, the components of a "generic" cochlear implant are as follows. A small head-worn microphone transduces the acoustic stimulus to an electrical signal. This electrical signal is then sent to a speech-processing unit. This unit typically extracts certain relevant aspects of the speech signal, such as the fundamental frequency and the formant frequencies. The processing unit then sends an electrical code of this information to an external receiver worn behind the ear. The external receiver transmits this coded information across the skin to a surgically implanted internal receiver. The electrical output of the internal receiver is then distributed to the auditory nerve fibers in the inner ear via one or more electrodes that have been surgically inserted into the inner ear.

A wide variety of implant types have been investigated over the years with most fitting into a classification scheme based on the number of channels and number of electrodes. The number of channels refers to the number of parallel channels in the processing of the encoded information and the number of electrodes refers to the number of independent electrical connections carrying that information to the nerve fibers of the inner ear. In general, the research conducted on cochlear implants over the past decade has indicated that the multiple-channel, multiple-electrode devices are the most beneficial. Although there are some individuals that have demonstrated truly remarkable performance with the implants, such as understanding conversational speech on the telephone, the vast majority of implantees appear to primarily receive timing information that is a valuable supplement to lipreading (visual information). Although the implant often does not appear to present users with true auditory sensation, it does make them aware of sound. Sound awareness and supplementation of visual information represent the two most common improvements experienced by implant users. Continued research in this area, however, is likely to bring additional improvements in performance.

At present, the primary candidates for a cochlear implant are those individuals who have acquired a profound bilateral hearing loss after they have developed speech and language and who receive only limited benefit from other prosthetic devices, such as the conventional hearing aid. This represents a rather small segment of the hearing-impaired population.

3.5 Rehabilitation for Persons with Tinnitus

Concluding this section on rehabilitative audiology, it should be noted that the aural rehabilitation process is not always centered on the loss of hearing. One of the most troubling concomitants of sensorineural hearing loss is the presence of tinnitus. Tinnitus is a "ringing," "whistling," or "buzzing" sensation perceived by the listener in the absence of sound. There appear to be many causes of tinnitus, and a wide variety of treatments have been suggested.[10] Head-worn noise generators that physically resemble hearing aids, for example, are often recommended for tinnitus sufferers. These noise generators essentially bring relief to the tinnitus sufferer by masking the tinnitus with the external noise. Frequently, residual masking effects are observed in which the tinnitus is reduced or eliminated for prolonged periods of time after removal of the tinnitus masker. As mentioned, there are a wide variety of causes and treatments for tinnitus, and the audiologist will frequently play a role in the rehabilitation of the tinnitus sufferer.

4 SUMMARY

This chapter has provided a very brief overview of the field of clinical audiology. As mentioned, audiologists are the primary professionals involved in the rehabilitation of the hearing impaired. Initial steps in the rehabilitation process involve the determination of the extent and severity of hearing loss and the site of lesion. Once this information is obtained, the next step typically involves the selection and evaluation of an appropriate prosthetic device for the hearing-impaired individual. Devices available include conventional hearing aids, assistive listening devices, vibrotactile aids, and cochlear implants. Conventional hearing aids and assistive listening devices are beneficial to the vast majority of hearing-impaired individuals, whereas vibrotactile devices and cochlear implants are primarily reserved for those profoundly impaired individuals unable to benefit from conventional amplification. Following the selection and evaluation of the appropriate device, additional aural rehabilitation, such as training in lipreading, auditory training, and training in effective communication strategies, is often provided by the audiologist.

This brief overview provides little more than a thumbnail sketch of clinical audiology. Additional details regarding the many facets of clinical audiology can be found elsewhere.[11,12]

REFERENCES

1. A. J. Moss and V. L. Parsons, “Current Estimates from the National Health Institute Survey, United States, 1985.” *Vital and Health Statistics*, Series 10, No. 160, DHHS pub. no. (PHS) 86-1588, Public Health Services, Washington, DC, 1986.
2. U.S. Department of Health and Human Services, “Aging America: Trends and Projections, 1987–1988 Edition,” U.S. Government Printing Office, Washington, DC, 1988.
3. *Specification for Audiometers*, ANSI S3.6-1989. Acoustical Society of America, New York, 1989.
4. J. W. Hall, *Handbook of Auditory Evoked Responses*, Allyn & Bacon, Boston, 1992.
5. J. T. Jacobson and J. L. Northern, *Diagnostic Audiology*, Pro-Ed, Austin, TX, 1991.
6. L. Braida, N. Durlach, R. Lippmann, B. Hicks, W. Rabinowitz, and C. Reed, “Hearing Aids—a Review of Past Research on Linear Amplification, Amplitude Compression and Frequency Lowering,” Asha Monograph No. 19, American Speech-Language-Hearing Association, Rockville, MD.
7. G. Walker and H. Dillon, “Compression in Hearing Aids: An Analysis, a Review and Some Recommendations,” National Acoustic Laboratories Rept. No. 90, Australian Government Printing Service, Canberra, 1982.
8. C. L. Compton, *Assistive Devices: Doorways to Independence*, Assistive Devices Center, Gallaudet University, Washington, DC, 1989.
9. Working Group 95, Committee on Hearing, Bioacoustics and Biomechanics (CHABA), National Research Council, “Personal Speech Perception Aids for the Hearing Impaired: Current Status and Needed Research,” *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 637–685.
10. D. McFadden, *Tinnitus: Facts, Theories and Treatments*, National Academy, Washington, DC, 1982.
11. F. H. Bess and L. E. Humes, *Audiology, the Fundamentals*, Williams & Wilkins, Baltimore, MD, 1990.
12. J. Katz (Ed.), *Handbook of Clinical Audiology*, 4th ed., Williams & Wilkins, Baltimore, MD, 1994.

121

PERCEPTION OF COMPLEX WAVEFORMS

CHARLES S. WATSON AND GARY R. KIDD

> *The most important things to know about a sound are not its frequency, amplitude, and phase, but the direction and distance of its source, and whether that source is something that you can eat, or that can eat you.*
>
> —Lloyd Jeffress
>
> Introduction to his graduate seminar on *Hearing: Building Blocks for Model Builders* (1963)

1 INTRODUCTION

Listeners' abilities to hear the details of complex sounds and to recognize the overall spectrotemporal form, or structure, of such sounds, have been the subject of considerable study in the past 15–20 years.[1–3] Much of this work reflects a growing interest among psychoacousticians in waveforms that have more in common with the familiar sounds of objects and events in our environment than do the pure tones and Gaussian noises that dominated psychoacoustical research years ago. This has led to a greater appreciation of the influence of the larger spectral and temporal context on the perception of the components of complex sounds and on the perceived relations among the components. In this chapter, we first define the category of "complex sounds" and then review research on several subcategories of complex sounds that have already received a good deal of attention. Next we discuss two related issues, namely perceptual learning and stimulus uncertainty, that are crucial for an understanding of listeners' abilities to discriminate and identify complex sounds. Finally, we review some practical considerations for the selection of sounds to be used as warnings or informative signals.

1.1 Definition of Complex Sounds

The term *complex sound* is used quite generally to refer to any sound that is more complex than a pure (sinusoidal) tone. Obviously, such a definition has little taxonomic value. It is used among psychoacousticians largely as an acknowledgment that research with tones and stochastic noises (which dominated the field until fairly recently) failed to address many crucial questions about human hearing. The research with these "simple" sounds has been quite fruitful, especially in yielding information about the peripheral auditory system, but it has become clear that a more complete account of human hearing requires research with complex sounds and with new tasks designed to reveal the role of central (or cognitive) mechanisms as well as to provide new insights into peripheral mechanisms. Thus, research with complex sounds is more likely to address issues of attention, encoding strategies, and perceptual organization. Some of these issues will be discussed in the following sections.

For the purposes of this chapter, we will use the "anything-but-tones-or-stochastic-noises" definition and add to it merely the obvious subcategories of complexity in either the temporal, spectral, or spectral *and* temporal domains. A recent report from the Committee on Hearing, Bioacoustics, and Biomechanics (CHABA)[2] on the classification of complex sounds came to roughly this same nondefinition but added to it the criterion that the sounds to be considered would be brief. This restriction is accomplished by default, since most research on complex sounds has been concerned with waveforms lasting a few seconds or less.

Some special categories of complex sounds have, of course, long been singled out for attention: those classes of acoustic events that are logically considered together because of their common sources, or functions (pur-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

poses), especially the sounds of speech and of music, but also the sounds of human hearts, of things that move in the ocean, and sounds designed to encode messages (e.g., Morse code, whistle languages). The topics of speech and music perception are covered in other chapters in this volume. Apart from these two areas, most recent research on complex sounds has been mainly concerned with the hearing of sounds whose "common sources" are the waveform generators and digital synthesizers of the research laboratory, rather than real-world sounds, such as a cavitating propeller or a radiotelegrapher's "bug." If science progresses in its normal manner, it seems likely that our understanding of the hearing of these real-world sounds will eventually be enhanced by what is learned from controlled studies of laboratory-generated sounds with various forms of complexity.

2 CATEGORIES OF COMPLEX SOUNDS

A practical categorization scheme for complex sounds groups them according to whether the primary structural complexity is spectral, temporal, or spectrotemporal (i.e., an orderly covariation in spectral and temporal structure). Since all complex sounds can be constructed by the combination of sinusoidal components, spectral structure can be manipulated by varying the number, frequency, intensity, and phase relations of the components of a complex waveform. The spectral content of a sound may change over time, either by varying the values of each of these attributes or by presenting each component sequentially to create a sequence of tones. When spectral structure changes over time, it is useful to consider the spectral and temporal changes together as defining a spectrotemporal structure. In some cases, the spectral structure remains fixed and the durations of a sequence of identical sounds are manipulated, so that the sounds are identifiable primarily by their temporal structure.

For any type of complex sound, a listener can be asked to perform the fundamental tasks of detection, discrimination, and identification, just as with simple sounds. However, new issues arise with complex sounds. Some tasks require the listener to resolve particular details of sounds while others require them to attend to more holistic properties (such as the pitch contour or the overall temporal structure). As the number and complexity of stimulus dimensions increase, additional questions arise about listeners' abilities to attend to information in these multiple dimensions.[4–6] Also, the detection of complex sounds (especially in the presence of other sounds) is complicated by the need to detect relations among components that are spectrally and/or temporally separated. This has led to an interest in identifying factors that make certain spectrotemporal components (or configurations of components) more salient than others.

In the following sections, a variety of perceptual tasks will be discussed in the context of each of the categories of complex sounds defined above. This brief overview covers several basic aspects of complex sound perception. Although research with speech and music is not considered in this chapter, many of the issues discussed are quite applicable to those special classes of sounds.

2.1 Spectrally Complex Sounds

This section is concerned with sounds for which complexity is primarily in the spectral structure. The role of temporal structure is typically downplayed in research with such sounds by holding timing parameters constant at values that are considered to be simple, representative, or in some way appropriate for the problem under investigation.

Simultaneous Complexity

Detection and Discrimination of the Components of Complex Sounds. If the components of a complex sound are considered to be the individual sinusoids from which the sound can be synthesized, then we may ask about listeners' abilities to resolve the properties of those sinusoidal components or "partials" of complex sounds under various conditions. Although several harmonically related simultaneous sinusoids are generally heard as a single sound, listeners are able to hear the first five to eight partials in a harmonic series or in an inharmonic series created by slight mistunings of the partials of a harmonic series.[7] A major factor appears to be the critical bandwidth within which the partials occur. Partials that fall within one critical bandwidth are difficult to hear individually. However, other factors also appear to play a role, especially for partials lower than 1000 Hz. For example, decreasing the number of partials can make it possible to distinguish partials at smaller frequency separations[8] (although this is not always the case[9]). Also, it has been reported that musicians can discern individual partials with considerably smaller frequency separations than nonmusicians, suggesting individual differences in frequency-resolving power *or* that auditory experience may play a role[10].

The ability to detect a change in the frequency of a partial does not necessarily imply that the listener will report "hearing out" that partial. In a harmonic series, small changes in the frequency of a partial are sometimes described as a pitch shift (of the sound as a whole), a change in timbre, or both. Larger changes in frequency (roughly 3–6%) can result in the altered partial being perceptually segregated from the other partials in the complex.[11] It thus appears that perceptual grouping mechanisms and other mechanisms sensitive to relations among partials become more important than simple frequency-resolving power for the discrimination of changes in the individual partials of some complex sounds.

There are many other cases in which listeners are able to detect a change in one partial or in one property of a complex sound by detecting some other property of the sound that changes as a result of alterations in the relations among partials. For example, a change in spectral density in sounds with many partials changes the spectrum, but it can also introduce temporal fluctuations in the level of the sound that, at least in some cases, are more salient than the spectral change.[12]

In an extensive series of experiments, Green and his colleagues have investigated the detection of changes in the amplitude of a single sinusoidal component in complex sounds consisting of several simultaneous components.[13] This work has shown that the relative amplitudes of a complex (i.e., its spectral profile) can be discriminated quite well despite random variation in the overall level of the complex and despite random variation in the phase relations among the components (which alter the waveform but leave the spectral profile unchanged). This supports the principle that it is the relational properties of complex sounds that are important for sound identification.

Detection and Identification of Complex Signals in Complex Backgrounds. When listeners are asked to detect or identify a target sound in the presence of other complex sounds, it is often found that the spectral and temporal relations among components of target and background sounds have a strong influence on performance. One of the simplest examples involves the detection of a target tone centered in narrow-band noise when flanking bands of modulated noise are also present. A common finding has been that when the temporal envelopes of the on-target and flanking bands of noise are correlated, there is a reduction in the masking normally observed when the temporal envelopes are not correlated. This type of release from masking has been labeled comodulation masking release (CMR).[14] Many aspects of CMR have been investigated[15] (see Chapter 115); among the most important findings are (1) masking release occurs when flanking bands are presented nearly one octave above or below the signal frequency (well outside of a critical band centered on the signal frequency) and (2) masking release occurs even when the comodulated flanking bands are presented to one ear and the target signal and on-target noise are presented to the other ear. These results suggest at least a portion of the CMR involves a type of central mechanism that operates across critical bands.

Several mechanisms have been suggested to explain CMR (see Chapter 115). Perhaps the most general explanation invokes the mechanisms of auditory scene analysis.[16,17] It has been suggested[18] that a perceptual grouping mechanism that automatically groups spectral components with common temporal envelopes may be responsible for CMR as well the related phenomena of comodulation detection difference (CDD) and modulation detection interference (MDI). In both CDD and MDI, comodulated sounds interfere with detection or resolving power for target sounds. In CDD, detection thresholds for a modulated band of noise (target) are higher when the envelope of a flanking band of noise is correlated with the target band of noise than when they are not correlated.[19,20] In MDI, thresholds for the detection of amplitude modulation of a tonal carrier are higher when a flanking band of noise is modulated at the same rate.[18] Thus it appears that a grouping mechanism that causes components with correlated envelopes to be heard as a single sound may facilitate the detection of a tone in modulated noise in the CMR case, because the tone is perceptually segregated from the noise bands heard as a single (and separate) auditory object. Similarly, in the case of CDD and MDI, the signal may be more difficult to discern because it becomes fused with the comodulated flanking noise.

Correlated changes in amplitude among spectral components can also facilitate selective attention in more natural listening situations, where listeners attempt to attend to particular components of complex waveforms or are required to detect or identify a sound (often associated with a particular source) that is presented either in noise or along with other sounds. Many instances of this have been demonstrated by McAdams[16] and Bregman.[17] Amplitude modulation is present to some degree in many naturally occurring sounds, and the auditory system may have evolved to take advantage of this property to help segregate sources in the auditory environment.

Several other stimulus properties have been shown to influence the perceptual grouping of components of complex sounds and thereby facilitate selective attention. In general, any stimulus property that has distinct values for different sets of components may provide a basis for selective attention. Frequency modulation can be as effective as amplitude modulation in facilitating the segregation of components.[16,21,22] When a difference in the onset and offset times of components is present, components that are not closely synchronized tend to form separate groups.[23,24] Differences in frequency range,[7,25] intensity,[26–28] and spatial location among sets of components[29–31] can also promote segregation.

There are bases for perceptual grouping of spectral components in addition to the common properties shared by a set of components. The frequency relations among a set of components can be a powerful determinant of the perceptual organization of complex sounds. Sets of spectral components that are harmonically related are generally perceived as one sound. This is an ecologically valid proclivity since many naturally occurring sounds, including speech and animal vocalization, consist of such sets of harmonics.

Sequential Complexity When the components of a complex sound are arranged sequentially in time, a different set of factors determines the discriminability of changes in the components. With these sounds, listeners must focus attention in time as well as in the spectral dimension. Consequently, both temporal and spectral relations among pattern components influence pattern discriminability.

Random Patterns. Watson and his colleagues have studied several aspects of the discriminability of changes in randomly generated sequences of tones.[32,33] Much of the early work employed isochronous (i.e., constant inter-onset-interval) 10-tone sequences with tone durations of 40 ms and no pauses between the tones. These sequences were intended to serve as very crude nonspeech analogs to words, with tone durations approximately equal to the shortest phonemes in rapid conversational speech. Typically, the change to be detected within a pattern was an increment in the frequency (or, in some cases, the duration or level) of a single "target" tone. More recently a broader range of component and pattern durations (including anisochronous patterns) have been studied, and many of the same principles of auditory pattern perception have been found to apply. The most important of those principles are summarized below.

Stimulus uncertainty (discussed in more detail in Section 4) is one of the most important factors in the perception of tonal patterns. It appears that for a wide range of pattern parameters, given enough repetitions of a single pattern with a constant target-tone position, listeners can learn to focus attention equally well at most spectrotemporal locations, regardless of the details of the pattern structure. Therefore, it is important to keep in mind that the factors described below are relevant only under high stimulus uncertainty (defined operationally by drawing stimuli at random from a large set on a trial-to-trial basis).

When listeners are asked to detect a change in the frequency of a single (target) component in a 10-tone pattern, considerable variation in discrimination performance is accounted for by the spectral and temporal position of the altered tone. Later occurring and higher frequency components are generally resolved with much greater accuracy than the other components. Listeners require as much as a 50–60% change for the early, low-frequency targets and as little as a 2–3% change for the later, high-frequency targets.[34]

Another factor that plays a major role in determining resolving power for component changes is the proportional duration of the target tone. It has been demonstrated that each individual component of an unfamiliar sequence of tones is resolved with an accuracy that is a function of its proportion of the total duration of the sequence. This *proportion-of-the-total-duration* (PTD) principle has been shown to apply over a wide range of total pattern durations, numbers of components, and component durations.[35,36] Comparable results have been obtained in recent studies that have shown that the detectability of a segment of "frozen" Gaussian noise from one sample embedded within a longer context of random noise depends on the proportion of the total duration occupied by the frozen-noise segment, rather than on its absolute duration.[37] Thus, it appears likely that the PTD principle applies to unfamiliar complex sounds in general rather than being limited to sequences of tones. There has been some debate as to whether the PTD rule reflects the effects of the relative timing of tones within a given pattern on auditory attention[38] or the effect of the variance within a set of tones used in a given experimental context.[39,40]

Pattern Structure and Perceptual Organization. Other factors that influence pattern discriminability involve the effects of the details of tone relations within patterns that are often difficult to assess in research with randomly generated sequences. Much of the work examining the effects of pattern details addresses proclivities rather than capabilities.[41] That is, the responses are subjective reports that cannot be scored as correct or incorrect, and their relation to auditory capabilities is unclear. For example, much of the literature on stream segregation and auditory perceptual organization requires listeners to report changes in the way a pattern "sounds" as frequency relations are altered. Although there is good reason to believe that many changes in perceived pattern organization are correlated with changes in capabilities (e.g., when certain tones are judged to be more salient, it seems likely that their properties will be resolved more accurately), these correlations have rarely been tested.

An example of a relation between proclivities and capabilities was shown in studies of auditory stream segregation in sequential patterns. In the context of sequential patterns, auditory stream segregation (or "streaming") refers to a perceptual splitting apart of a sequence of tones into two or more separate, but concurrent, sequences (or "streams"). Subjectively, it is difficult to attend to more than one stream at a time, and between-stream frequency relations are much more difficult to attend to than frequency relations within a stream. Bregman and Campbell[42] demonstrated that this perceived independence of streams is also associated with a degraded ability to judge the temporal order of tones located in different streams compared to tones in the same stream. However, there is still considerable debate as to the extent of this degradation and the conditions under which it occurs. (See Ref. 17, pp. 143–155, for a summary of views on this issue.)

Apart from the possible relations between proclivities and capabilities, there has been considerable interest in stream segregation because of its implications for auditory attention and pattern identification. A major focus has been that of identifying the conditions under which streaming occurs. Although the precise conditions for streaming vary considerably with different pattern and task conditions, it is clear that streaming becomes more likely with increases in the frequency separation between subsets of tones in a pattern. Streaming based on differences in spatial location,[43,44] timbre,[45,46] and intensity[47] has also been demonstrated (although intensity-based streaming is relatively weak). Temporal relations are also important in the formation of streams. The frequency separation required for the formation of streams generally decreases as the temporal separation increases. This has led to speculation that there may be a limit on the rate of frequency change a listener can track in patterns of discrete tones.[48,49] However, estimates of this rate vary considerably as a function of pattern context, pattern repetition, and various attentional factors. More complex aspects of temporal structure and its relation to spectral structure also affect stream segregation and are discussed in more detail in the section on spectrotemporal complexity.

Many of the factors that affect auditory stream segregation are summarized by Gestalt principles of organization (see Ref. 17 for a review). Principles such as similarity, good continuation, and common fate play a role in determining the perceived organization of auditory patterns. These factors also affect the salience of particular components within patterns and thus may affect the likelihood of detecting changes in the properties of those components.

2.2 Temporally Complex Sounds

For many complex sounds the primary information that they carry is in the temporal domain. For example, the difference between the sounds of a breaking bottle and a bouncing bottle, the difference between the sounds of walking and running, and the meaning of Morse code sequences are conveyed primarily or exclusively via temporal structure. There has been considerable interest in various aspects of the perception of temporal patterns. Much of the work has focused on the ability to recognize or discriminate temporal patterns. Several experiments have shown that changes in a tone duration or intertone interval can be detected about as well in the context of a temporal pattern as they can when presented in isolation,[50–53] at least when tested under low-uncertainty conditions (see Section 4). Discrimination performance has typically been found to conform to Weber's law. That is, discrimination thresholds are approximately equal in terms of the proportional change in duration ($\Delta t/t$) over a wide range of base durations. Measured threshold values of $\Delta t/t$ range from approximately 0.05 to 0.20, depending on the task and stimulus properties. Experiments with isolated temporal intervals (i.e., not embedded within multitone patterns) have generally obtained Weber ratios near 0.10 for intervals between about 200 and 2000 ms.[54–57]

Although listeners can discriminate duration changes in multicomponent temporal patterns about as well as with isolated intervals under certain conditions, there are several properties of temporal patterns that have systematic effects on discrimination performance. For example, at intertone intervals less than 100 ms or so, temporal discrimination tends to be somewhat better for intervals nearer the end of the sequence. Also, decreases in the tempo of a pattern not only make temporal discrimination less difficult (as intervals are increased to 100 ms or more) but also attenuate the effects of pattern position.[52,58]

It has also been shown that temporal accents (produced by the pattern of relative durations of tones and/or pauses) affect the perception of temporal pattern structure.[59,60] While temporal accents affect the recognition and reproduction of melodies and rhythms,[58–60] data from some temporal discrimination experiments suggest that discrimination performance can be better accounted for by a model based on Weber's law than by many of the more complex accent-based models.[58]

Sorkin[53,62] has proposed a pattern correlation model to account for the discrimination of temporal patterns. According to the model, discrimination is performed by computing the correlation between the sequence of temporal intervals within two patterns to be compared. This model provides a good account of discrimination performance under a variety of conditions, including different mean interval durations and different pattern lengths (i.e., numbers of intervals). However, the model cannot account for many effects based on similarity in the relative timing or accent structures since a correlation will not always reflect this type of similarity. For example, the model does not discriminate between changes in pattern rate that preserve the pattern of relative durations (i.e., multiplicative expansions or contractions of intervals) and ones that do not (e.g., additive expansions or contractions) as long as the pattern correlation is the same (see Ref. 62). While this is a problem for the perception of rhythmic patterns and many types of rule-governed temporal patterns, Sorkin's model may be appropriate for the perception of random, arrhythmic patterns.

The ability to detect timing changes in randomly generated patterns of temporal intervals that have undergone tempo transformations (i.e., relative-time-preserving expansions or contractions of durations) has been studied by Sorkin and Montgomery[62] and by Kidd and Watson.[63] Although it is fairly easy to identify regular, music-like rhythms despite fairly large changes in tempo (within limits), listeners had considerable difficulty determining whether tempo-altered comparison patterns were different from standard patterns in these studies. However, the greater difficulty with the random-timing patterns appears to have little to do with inherent properties of the temporal structure. Both of the studies with random patterns also used a different random pattern on every trial. When Kidd and Watson tested subjects under minimal uncertainty (i.e., a single pattern and target position), they found that subjects were affected very little by tempo changes and could detect changes in the relative duration of components about as well as listeners can detect changes in absolute duration ($\Delta t/t \approx 0.10$).

Another aspect of the perception of temporal properties has been shown in studies of the effects of the larger temporal context on the perception of durations of individual events within patterns. The pattern of component durations within a complex sound can set up expectations in the listener for relatively long or short durations at specific temporal locations. When event durations deviate from these expectations, their perceived duration is often altered in the direction that the stimulus departs from the expectation (i.e., a perceptual contrast effect).[59, 64]

Discrimination and Identification of the Temporal Order of Components of Complex Sounds Green[65] summarized convincing evidence that the order of spectral elements with a complex waveform can be discriminated for waveforms as brief as 1–2 ms. At the other extreme, Warren et al.[66] have shown that listeners often are unable to identify the order of a few familiar sounds (excluding speech and music) presented in repeating sequences, unless the individual sounds are longer than 200 ms. Hirsh[67] found that the onsets of tones, or of noise bursts differing in pitch, must be separated by 18–20 ms for listeners to identify their order of occurrence. It appears that this wide range of results can be explained in terms of two principles. The very brief durations required for discrimination between two sequences differing only in the order of spectral elements are obtained when the discrimination tasks permit comparison of the immediate sensory memories (or "traces"; see later section on cognitive theories) elicited by the two waveforms. The experiments yielding longer estimates of the durations required for order identification generally require that the listener compare a presented sequence to some form of generalized memory for that sequence, or "sensory prototype." Finally, as Warren[68] has proposed, correct identification of the order of sounds in a continuing sequence must require the storage and mental "shuffling" of a set of verbal labels, in addition to comparison with stored prototypes. It seems likely, given the maximum temporal acuity described by Green and others,[65, 69] that listeners may generally be expected to be able to learn to make use of temporal order information for any stimuli for which experience has allowed the development of the necessary prototypes. Hirsh's results, and others, suggest that the generalized prototype for "auditory successiveness" that most listeners use in trying to label the order of unfamiliar stimuli is one for which stimulus onsets are no less than 18–20 ms apart, while smaller onset differences are closer to their prototypes for simultaneity.

2.3 Spectrotemporal Complexity

All sounds have structure in both the spectral and temporal domains. When we try to study one dimension alone by simplifying the structure on the other, we run the risk of missing important interactions between the two dimensions. The significance of these interactions is evident in the following summary of research on complex sounds with spectral and temporal complexity. Many other examples of these interactions are found in speech and music, as described in other chapters in this volume.

Sequences of Tones Over the past several years, Jones and her colleagues have investigated effects of spectral and temporal structure on the perception of temporal patterns (see Refs. 48, 59, 61 and 70 for reviews of data and theory). Much of this research has shown that the relative timing of events within a pattern can have a strong influence on the way that auditory patterns are perceived, presumably through the influence of pattern timing on auditory attention. For example, events that occur at temporally predictable or accented temporal locations are detected and remembered more accurately than less predictable events.[71] The same attentional mechanism appears to be responsible for effects of temporal structure on the ability to judge the temporal order of tones within a sequence. The temporal order of tones within a pattern is more easily recognized when their timing is inconsistent with the rhythm of the other tones in the pattern than when the entire sequence is presented in an isochronous rhythm.[72] Listeners are apparently able to use the rhythmic structure of a pattern to direct their attention to specific tones when their timing is inconsistent with the temporal structure of the rest of the pattern.

Many of the effects of temporal structure on the perception of spectral properties of patterns highlight the intimate relation between spectral and temporal structure in defining a pattern's perceptual identity. For example, it is often very difficult to recognize a sequence of tones when its temporal structure is altered.[73,74] Also, changes in the coincidence of accents created by temporal structure and those created by spectral structure can make a sequence of tones difficult to recognize.[59] In many cases, timing may affect pattern discrimination through its effect on the perception of auditory streaming as discussed earlier. A change in the pattern of streaming alters the pattern's perceptual identity, and because relations between tones that form different streams are more difficult to discern,[42] a different set of pitch relations is heard. Of course, not all changes in pattern timing will affect listeners' abilities to recognize a sequence of tones or to detect a change in a sequence, but it is important to appreciate that many types of temporal changes will have large effects on these abilities.

The influence of spectrotemporal structure can also be seen in the effects of spectral structure on the perception of *temporal* properties of patterns. Frequency structure can affect the resolution of temporal relations through its effect on stream segregation. The arrangement of large and small frequency changes within a pattern affects the formation of streams, and temporal relations between tones in separate streams are more difficult to resolve than relations within streams.[42,75] Also, high uncertainty with respect to the pattern of frequencies has been shown to be associated with degraded temporal resolution.[51,75]

Changes in Spectral Shape over Time For sounds that consist of several simultaneous components, spectrotemporal structure refers to changes in spectral shape over time. This can mean the addition or deletion of components or a change in the amplitude of components over time. The sounds of music, speech, and most environmental sounds that we experience have this type of structure. There are several examples in each of these categories that show that it is the pattern of change in spectral structure over time that identifies these events. In studies of musical timbre it has been observed that the information required for the identification of musical instruments is substantially degraded when the attack and decay transients are removed and only the steady-state portion of the waveform is presented.[76,77] Similar effects have been demonstrated in speaker identification[78,79] and vowel identification.[80–82]

There has been considerably less work on the perception of environmental sounds. Some of this work has been done in applied settings where researchers have been concerned with such issues as the use of sound to facilitate human–computer interaction[83] and the manipulation of "sound quality" to improve the appeal of sound-emitting products.[84] Much of the basic research in this area has been concerned with identifying the properties of the waveform that allow a listener to identify the object or event (or specific properties of the event) that produced the waveform. Work has been done on the perception of breaking and bouncing bottles,[85] hand claps,[86] footsteps,[87] banging pots and pans,[88] and a variety of other common sounds.[89] (See Refs. 90–92 for discussions of research on auditory event perception.) In general, listeners have been found to be quite accurate at identifying many events and event properties by way of auditory information. However, little is known about which particular properties of the waveforms are important in identifying various types of sounds. The work that has been done suggests that, as with speech and musical instruments, the critical information is often dynamic in nature. That is, the manner of change in spectral shape over time is often more informative than the steady-state frequency relations at any given time.

3 AUDITORY PERCEPTUAL LEARNING

Auditory perceptual learning (or training) is one of the dominant factors that determine listeners' abilities to detect, discriminate between, and identify complex sounds, perhaps second only to auditory sensitivity as measured by the audiogram. Unfortunately, in the field of psychoacoustics the overwhelming importance of the level of learning (or amount of training or of "familiarity," as the results of informal training tend to be called) has generally been regarded as a nuisance, rather than as a topic of research per se. Since the earliest systematic studies of human hearing, research reports described the training of subjects in terms such as "well-trained listeners were used" or "training was continued until no further improvement was noticeable." Only very rarely did authors describe their training procedures in detail, and even less often were data presented showing the changes in performance at various times during the course of the total period of training. In reviewing the scarce data that have been reported, however, it is possible to offer some preliminary conclusions about the time course of auditory perceptual learning.[93,94] Table 1 is organized in terms of the three primary questions that may be asked about listeners' abilities to hear various sounds, their abilities to detect, discriminate, and identify (columns), and the complexity of the stimuli, ranging from isolated pure tones to tonal patterns and Morse code (rows).[94] The entries are from a sampling of the few studies that have explicitly reported the time required to approach asymptotic performance. Two very broad generalizations are illustrated. The first is that the time required to approach asymptotic performance in an auditory psychophysical task can range from less than 1 h to several months (more than a year in some cases). The second is that the more information that is transmitted in a particular auditory task, the longer the time course of learning for that task. In other words, less information is transmitted by detecting the presence of a particular sound than by identifying it, but identifying one sound that occurred from a small set of potential signals provides less information than identifying one from a large set—and complex stimuli typically come in larger sets than simpler stimuli. These generalizations are not much more than casual observations based on examination of the data in Table 1, and they certainly deserve more systematic study. The important observation is that it can take so long to arrive at the level of performance of which the human is capable that casual listening to a set of potential signals may provide virtually *no* useful information about the way a trained listener may be able to respond to them. Despite its having been published in many other places, it is worthwhile to include Figure 1 (from Bryan and Harter.) Bryan (later President of Indiana University) was a first-generation experimental psychologist, and it is interesting that his study of the time course of learning to receive and send Morse code remains one of the most extensive reports of auditory perceptual learning ever published.[106] The great majority of modern studies of auditory learning would be complete within the first 2–3 weeks of Bryan and Harter's 48-week experiment (their subjects worked 6–8 h/day). Figure 1 shows no evidence of asymptotic performance after 48 weeks. This suggests that modern research has probably provided accurate estimates of human capability from studies of the detection or discrimination of simple sounds but that we may have greatly underestimated abilities to learn to identify complex sounds. One practical application of this conclusion is to the evaluation of systems for recoding speech, as in frequency-lowering electroacoustic hearing aids or aids that otherwise translate the sounds of speech to make them maximally intelligible to an impaired listener or other speech aids, including cochlear implants or sensory substitution aids.[93,94] Use of each of these devices require that a person learn a new sensory code, and the time to learn *any* code is clearly long enough that individual codes could not be given sufficient testing to determine the performance that might eventually be achieved with them. Some means of estimating long-term asymptotic performance from a priori considerations or from preasymptotic performance is clearly needed.

4 STIMULUS UNCERTAINTY

A dimension that might usefully be added to Table 1 is that of stimulus uncertainty. The ability to "hear out" the critical details of a complex sound depends much more strongly on stimulus uncertainty (or certainty) than do detection and discrimination abilities for simple stimuli. When the frequency of an isolated tone is changed from trial to trial, its threshold of detectability and the ability to notice small changes in its frequency are both degraded, compared to determinations with a fixed-frequency tone, but by *very small amounts*. The detection threshold is elevated by 1–2 dB, and the frequency difference limen is reduced from 4–5 to 8–10 Hz (for frequencies near 1000 Hz). However, if the same tone is presented as a member of a multitone pattern, detection and discrimination performance for that "target" tone are very strongly determined by stimulus uncertainty.[34,32] Of course, there are several types of uncertainty. For example the entire pattern can be varied from trial to trial or the same pattern can be presented on each trial while the temporal position or frequency of the target tone varies over trials. At the highest levels of stimulus uncertainty, thresholds for the detection of tonal components of multitone patterns can be elevated by 30–50 dB, compared to the minimal-uncertainty case (same pattern on every trial, same component subject to change), and the detectability of changes in frequency, intensity, or duration is similarly elevated, as illustrated in Table 2.[32] The importance of estimating listeners' abilities under minimal-uncertainty conditions was under-

TABLE 1 Time Required for Auditory Perceptual Learning for Detection, Discrimination, and Identification and for Simple versus Complex Sounds

Stimuli	Detection	Discrimination	Identification
Single tones;	1.6 h[95]	4 h[96]	24–28 wks[97]
other simple	<1.0 h[98]	4 h[99]	20 h[100]
stimuli	<1 week[101]	7 h[102]	4 weeks[103]
Complex sounds;	<14 h[104]	6 h[105]	>40 weeks[106]
tonal sequences	<30 h[104]	>20 h[107]	>60 h[108]
		>36 h[51]	

Source: From Ref. 94.

TABLE 2 Detection and Discrimination Thresholds ($d' = 1.0$) for a 554-Hz, 40-ms Tone Presented in a Ten-Tone Pattern under Three Levels of Stimulus Uncertainty

Uncertainty Level	ΔF (Hz)	ΔT (ms)	ΔI (dB)	Detection (dB, SPL, $d' = 1$)
High	2400	56	9.9	58
Medium	49	27	6.8	46
Low	16	6.7	2.5	25

Source: From Ref. 32.

scored by an experiment in which minimal-uncertainty training on each of eight tonal patterns was followed by a condition in which those patterns were presented in random order.[28] The accurate discrimination achieved under minimal-uncertainty training was maintained in the random-order condition, suggesting that a relatively long-term ability to hear out the details of these familiar patterns had been established. If carryover of training to the more natural conditions did not occur, the extremely accurate discrimination performance achieved under minimal-uncertainty conditions would be little more than a laboratory curiosity.

5 AUDITORY WARNING SIGNALS

Several authors have reviewed the auditory research literature in terms of the guidance it can provide for such practical issues as the selection of optimally effective warning signals and sets of such signals.[109–112] While giving useful advice for the design of new signals and the evaluation of existing ones, these authors generally agree that too little is yet known about the hearing of complex sounds to provide detailed guidance. The only approach that has been successful is to select signals based on a few very general principles, as discussed in the following sections, and then train listeners intensely in their use.

5.1 Level Relative to Background Noise

The most obvious criterion for a warning signal is that it be audible. With a signal that is spectrally and temporally complex, audibility is determined by a combination of the power spectrum of the signal and that of the background noise in which it will be presented. As discussed in Chapter 115, auditory sensitivity is reasonably well modeled by an array of bandpass filters, each followed by a leaky integrator. To a first approximation, the level of a signal that is required for it to be audible above a masker is reached when the energy in a critical-band filter centered on the most prominent peak in the signal's spectrum is the same as that found in the masker in the same region. For midfrequency pure tones, this means that a tonal signal must be 15 dB or more above the spectrum level of a broadband masker. But in the case of a warning signal with a broad spectrum, it might mean that the threshold level for the warning signal is roughly the same as that of the masker within the critical band containing the most energy. Detailed procedures for estimating the threshold levels using filter shapes that closely approximate those of the human auditory system are given by Patterson.[110] Behaviorally measured thresholds are easy to obtain using the procedures described in Ref. 111. Empirical cross-checking seems advisable using the actual noise spectrum that will be present.

Having established the level required to make the warning signal audible, the next question is, how far above that level should it be presented? As Patterson and Milroy[113] and Sorkin[111] agree, 6–10 dB above threshold guarantees nearly perfect detection under laboratory conditions, but 16–22 dB may be needed to assure rapid responses in the field. When a warning signal exceeds 30 dB above threshold, its effects may actually be disruptive, including startle responses and the prevention of conversation about the condition being signaled.[114] Sorkin recommends that nonauditory channels (lights, tactile displays) be used when background noise would require warning signal levels in excess of 115 dB SPL.[111]

5.2 Spectral Patterns

The spectral properties of patterns influence their salience, or value as an alerting stimulus, although to a somewhat smaller degree than their temporal pattern (see following section). Causing the spectral pattern to change over time can also increase a signal's effectiveness, as demonstrated in emergency vehicle sirens ("wailers") used in heavy traffic situations. Multiple energy peaks yield a more distinctive pattern and one that is less likely to be masked. The manner in which the auditory system processes spectral components results in the perception of a fundamental pitch whenever several of its harmonics are present in a spectrum, thus making the fundamental pitch very resistant to masking. It

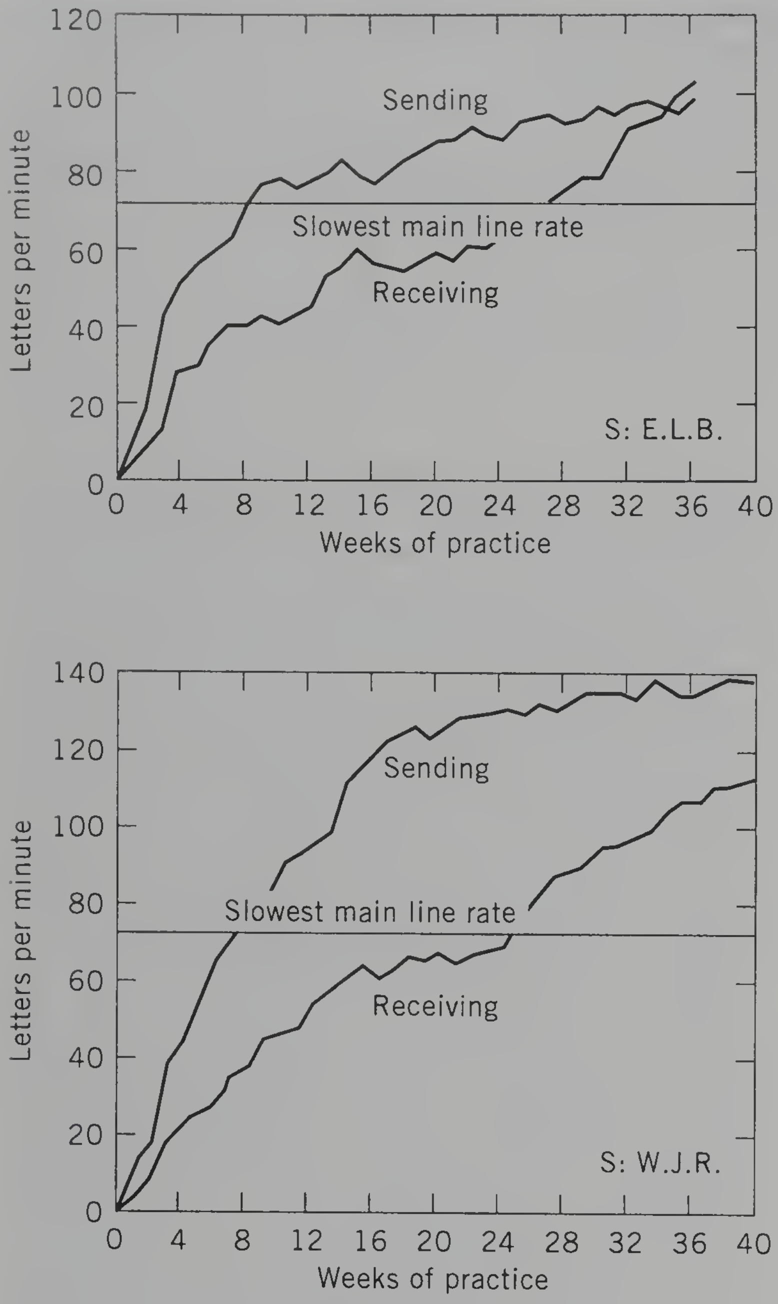


Fig. 1 Time course of learning Morse code for two subjects (as reported by Bryan and Harter[106]).

is also useful to recall, in the design of warning signals, that although masking spreads upward in frequency, high frequencies tend to be more attention demanding than do low.

5.3 Temporal Patterns

While there have been only a few studies of the matter, most authors concerned with the design of auditory "displays" have advised using a temporal pattern to achieve a distinctive signal. An international standard evacuation signal based solely on temporal characteristics was recommended by the Committee on Hearing, Bioacoustics, and Biomechanics (CHABA) of the National Research Council.[115]

5.4 Structure and Size of Sets of Warning Signals

Such auditory skills as the reception of Morse code by skilled operators or rapid "on-line" translation from one human language to another suggest that human listeners might be able to learn to use a very large array of auditory warning signals, with an equally varied catalog of meanings. However, several studies suggest that when limited by practical amounts of training, the maximum number of auditory warnings to which operators can respond appropriately may be rather small.

Patterson[97] taught groups of naive listeners to identify cockpit warning signals, employing 10 signals currently in use on civil aircraft. While listeners quickly learned to identify 4–6 of the signals, considerably longer times were required to achieve good performance on the entire set of 24 sequences. Similar results were obtained by Leek and Watson,[107] who trained listeners to identify 24 tonal patterns. The 12-tone patterns included the possible sequences of 4 mutually discriminable 3-tone "minipatterns." While listeners learned to identify the 4 minipatterns at better than 90% correct after 1–2 h of training, many more hours were required to learn to identify the entire set of 24 sequences. Based on this and other similar results, Patterson recommended small sets of warnings for which immediate action is required, possibly no more than six or seven.[114]

6 NEEDED RESEARCH

The human auditory system, and presumably that of most mammals, is remarkably sensitive and possessed of very accurate temporal acuity and at least moderate spectral acuity. But exactly how well listeners equipped with that system can learn to process arrays of complex sounds we simply do not know. Despite the large number of studies of the hearing of complex sounds in the past 20 years, most of that work has still been focused on the "hard-wired" properties of the peripheral auditory system. Much more must be learned about the role of the central auditory system in the perception of complex sounds, including topics such as the effects of the size of the stimulus catalog and effects of various types of structural regularities or "grammars." Studies of these topics are still clearly in their childhood, if not their infancy.

Acknowledgments

The discussion of auditory warning signals reflects useful conversations on that subject with Roy Patterson of the Medical Research Council, Applied Psychology Unit, Cambridge, U.K. Preparation of this chapter was partially supported by grants from the National Institutes of Health and from the Air Force Office of Scientific Research.

REFERENCES

1. W. A. Yost and C. S. Watson, *Auditory Processing of Complex Sounds*, Lawrence Erlbaum, Hillsdale, NJ, 1987.
2. Committee on Hearing, Bioacoustics, and Biomechanics (CHABA) of the National Research Council, *Classification of Complex Nonspeech Sounds*, National Academy Press, Washington, DC, 1989.
3. S. Handel, *Listening: An Introduction to the Perception of Auditory Events*, MIT Press, Cambridge, MA, 1989.
4. I. Pollack, "The Information of Elementary Auditory Displays," *J. Acoust. Soc. of Am.*, Vol. 24, 1952, pp. 745–749.
5. I. Pollack and L. Ficks, "Information of Elementary Multidimensional Auditory Displays," *J. Acoust. Soc. Am.*, Vol. 26, 1954, pp. 155–158.
6. G. R. Kidd and C. S. Watson, "The Perception of Complex Multidimensional Sounds," *J. Acoust. Soc. Am.*, Vol. 81, Suppl. 1, 1987, pp. S33 (abst.).
7. R. Plomp, "The Ear as a Frequency Analyzer," *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 1628–1636.
8. R. Plomb, *Aspects of Tone Sensation*, Academic Press, New York, 1976.
9. G. Vicario, "Some Observations in the Auditory Field," in J. Beck (Ed.), *Organization and Representation in Perception*, Erlbaum, Hillsdale, NJ, 1982.
10. D. R. Soderquist, "Frequency Analysis and the Critical Band," *Psychol. Sci.*, Vol. 21, 1970, pp. 117–119.
11. B. C. J. Moore, B. R. Glasberg, and R. W. Peters, "Relative Dominance of Individual Partials in Determining the Pitch of Complex Tones," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 1853–1860.
12. W. M. Hartman, S. McAdams, A. Gerzso, and P. Boulez, "Discrimination of Spectral Density," *J. Acoust. Soc. Am.*, Vol. 79, 1986, pp. 1915–1925.
13. D. M. Green, *Profile Analysis*, Oxford, London, 1988.
14. J. W. Hall, M. P. Haggard, and M. A. Fernandes, "Detection in Noise by Spectro-Temporal Pattern Analysis," *J. Acoust. Soc. of Am.*, Vol. 76, 1984, pp. 50–56.
15. J. W. Hall III, "Experiments on Comodulation Masking Release," in W. A. Yost and C. S. Watson (Eds.), *Auditory Processing of Complex Sounds*, Lawrence Erlbaum, Hillsdale, NJ, 1987.
16. S. McAdams, "Spectral Fusion, Spectral Parsing, and the Formation of Auditory Images," Ph.D. Dissertation, Stanford University, 1984.
17. A. S. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, MA, 1990.
18. W. A. Yost and S. Sheft, "Detecting Amplitude Modulation of Sinusoidal Carriers," *J. Acoust. Soc. Am.*, Suppl. 1, Vol. 83, p. S35 (abst.).
19. M. F. Cohen and E. D. Schubert, "The Effect of Cross-Spectrum Correlation on the Detectability of a Noise Band," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 721–723.
20. D. McFadden, "Comodulation Detection Differences Using Noise-Band Signals," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 1519–1527.
21. J. M. Chowning, "Computer Synthesis of the Singing Voice," in *Sound Generation in Winds, Strings, Computers*, Royal Swedish Academy of Music, Stockholm, 1980.
22. S. McAdams, "Segregation of Concurrent Sounds. I: Effects of Frequency Modulation Coherence," *J. Acoust. Soc. Am.*, Vol. 86, 1989, pp. 2148–2159.
23. A. S. Bregman and S. Pinker, "Auditory Streaming and the Building of Timbre," *Can. J. Psychol.*, Vol. 32, 1978, pp. 19–31.
24. R. A. Rasch, "The Perception of Simultaneous Notes. Such as in Polyphonic Music," *Acustica*, Vol. 40, 1978, pp. 21–33.
25. R. Plomp and A. M. Mimpin, "The Ear as a Frequency Analyzer II," *J. Acoust. Soc. Am.*, Vol. 43, 1968, pp. 764–767.
26. G. L. Dannenbring and A. S. Bregman, "Streaming vs. Fusion of Sinusoidal Components of Complex Waves," *Percept. Psychophys.*, Vol. 24, 1978, pp. 369–376.
27. J. P. Martens, "Comment on 'Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals.'" *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 626–628.
28. B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Thresholds for the Detection of Inharmonicity in Complex Tones," *J. Acoust. Soc. Am.*, Vol. 77, 1985, pp. 1861–1867.
29. E. C. Cherry, "Some Experiments on the Recognition of Speech with One and with Two Ears," *J. Acoust. Soc. Am.*, Vol. 25, 1953, pp. 975–979.
30. M. Kubovy and F. P. Howard, "Persistence of a Pitch-Segregating Echoic Memory," *J. Exper. Psychol.: Human Percept. Perform.*, Vol. 2, 1976, pp. 531–537.
31. D. R. Perrott and S. H. Barry, "Binaural Fusion," *J. Audit. Res.*, Vol. 3, 1969, pp. 263–269.
32. C. S. Watson, "Uncertainty, Informational Masking, and the Capacity of Immediate Recall," in W. A. Yost and C. S. Watson (Eds.), *Auditory Processing of Complex Sounds*. Lawrence Erlbaum, Hillsdale, NJ, 1987, pp. 267–277.
33. C. S. Watson and W. J. Kelly, "The Role of Stimulus Uncertainty in the Discrimination of Auditory Patterns," in D. J. Getty and J. H. Howard (Eds.), *Auditory and*

Visual Pattern Recognition, Lawrence Erlbaum, Hillsdale, NJ, 1981, pp. 37–59.

34. C. S. Watson, W. J. Kelly, and H. W. Wroton, "Factors in the Discrimination of Tonal Patterns. II. Selective Attention and Learning under Various Levels of Stimulus Uncertainty," *J. Acoust. Soc. Am.*, Vol. 60, 1976, pp. 1176–1186.
35. C. S. Watson, D. C. Foyle, and G. R. Kidd, "Limits of Auditory Pattern Discrimination for Patterns with Various Durations and Numbers of Components," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 2631–2638.
36. G. R. Kidd and C. S. Watson, "The "Proportion-of-the-Total-Duration Rule" for the Discrimination of Auditory Patterns," *J. Acoust. Soc. Am.*, Vol. 92, 1992, pp. 3109–3118.
37. S. F. Coble and D. E. Robinson, "Discriminability of Bursts of Reproducible Noise," *J. Acoust. Soc. Am.*, Vol. 92, 1992, pp. 2630–2635.
38. G. R. Kidd, "Proportional Duration and Proportional Variance as Factors in Auditory Pattern Discrimination," *J. Acoust. Soc. Am.*, Vol. 97, 1995, pp. 1335–1338.
39. R. A. Lutfi, "A Model of Auditory Pattern Analysis Based on Component-Relative-Entropy," *J. Acoust. Soc. Am.*, Vol. 94, 1993, pp. 748–758.
40. R. A. Lutfi, "Further Comments on Proportional Duration and Proportional Variance as Factors in Auditory Pattern Discrimination," *J. Acoust. Soc. Am.*, Vol. 97, 1995, pp. 1339–1340.
41. C. S. Watson, "Psychophysics," in B. B. Wolman (Ed.), *Handbook of General Psychology*, Prentice-Hall, Englewood Cliffs, NJ, 1973.
42. A. S. Bregman and J. Campbell, "Primary Auditory Stream Segregation and the Perception of Order in Rapid Sequences of Tones," *J. Exper. Psychol.*, Vol. 89, 1971, pp. 244–249.
43. W. J. Dowling, "Rhythmic Fission and Perceptual Organization," *J. Acoust. Soc. Am.*, Vol. 44, 1968, p. 369 (abst.).
44. W. J. Dowling, "The Perception of Interleaved Melodies," *Cogn. Psychol.*, Vol. 5, 1973, pp. 322–337.
45. D. L. Wessel, "Timbre Space as a Musical Control Structure," *Comput. Music J.*, Vol. 3, 1979, pp. 45–52.
46. P. Singh, "Perceptual Organization of Complex Tone Sequences: A Tradeoff between Pitch and Timbre," *J. Acoust. Soc. of Am.*, Vol. 82, 1987, pp. 886–899.
47. L. P. A. S. Van Noorden, "Minimum Differences of Level and Frequency for Perceptual Fission of Tone Sequences ABAB." *J. Acoust. Soc. Am.*, Vol. 61, 1977, pp. 1041–1045.
48. M. R. Jones, "Time, Our Lost Dimension: Toward a New Theory of Perception, Attention, and Memory," *Psychol. Rev.*, Vol. 83, 1976, pp. 323–355.
49. L. P. A. S. van Noorden, "Temporal Coherence in the Perception of Tone Sequences," Ph.D. Dissertation, Eindhoven University of Technology, The Netherlands, 1975.
50. A. R. Halpern and C. J. Darwin, "Duration Discrimination in a Series of Rhythmic Events," *Percept. Psychophys.*, Vol. 31, 1982, pp. 86–89.
51. B. Espinoza-Varas and C. S. Watson, "Temporal Discrimination for Single Components of Nonspeech Auditory Patterns," *J. Acoust. Soc. Am.*, Vol. 80, 1986, pp. 1685–1694.
52. I. J. Hirsh, C. B. Monahan, K. W. Grant, and P. G. Singh, "Studies in Auditory Timing: 1. Simple Patterns," *Percept. Psychophys.*, Vol. 47, 1990, pp. 215–226.
53. R. D. Sorkin, "Perception of Temporal Patterns Defined by Tonal Sequences," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 1695–1701.
54. S. Abel, "Duration Discrimination of Noise and Tone Bursts," *J. Acoust. Soc. Am.*, Vol. 51, 1972, pp. 1219–1223.
55. C. D. Creelman, "Human Discrimination of Auditory Duration," *J. Acoust. Soc. Am.*, Vol. 34, 1962, pp. 582–593.
56. A. M. Small and R. A. Campbell, "Temporal Differential Sensitivity for Auditory Stimuli," *Am. J. Psychol.*, Vol. 75, 1962, pp. 401–410.
57. H. Woodrow, "Time Perception," in S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, Wiley, New York, 1941, pp. 1224–1236.
58. C. B. Monahan and I. J. Hirsh, "Studies in Auditory Timing: 2. Rhythm Patterns," *Percept. Psychophys.*, Vol. 47, 1990, pp. 227–242.
59. M. R. Jones, "Dynamic Pattern Structure in Music: Recent Theory and Research," *Percept. Psycholphys.*, Vol. 41, 1987, pp. 621–634.
60. D. Povel and P. Essens, "Perception of Temporal Patterns," *Music Percept.*, Vol. 2, 1985, pp. 411–440.
61. M. R. Jones and M. Boltz, "Dynamic Attending and Responses to Time," *Psychol. Rev.*, Vol. 96, 1989, pp. 459–491.
62. R. D. Sorkin and D. A. Montgomery, "Effect of Time Compression and Expansion on the Discrimination of Tonal Patterns," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 846–857.
63. G. R. Kidd and C. S. Watson, "Detection of Relative Duration Changes in Tonal Patterns," *J. Acoust. Soc. Am.*, Vol. 88, Suppl. 1, 1990, p. S147 (abst.).
64. G. R. Kidd, "Articulatory-Rate Context Effects in Phoneme Identification," *J. Exper. Psychol.: Human Percept. Perform.*, Vol. 15, 1989, pp. 736–748.
65. D. M. Green, "Temporal Auditory Acuity," *Psychol. Rev.*, Vol. 78, 1971, pp. 540–551.
66. R. M. Warren, C. J. Obusek, R. M. Farmer, and R. P. Warren, "Auditory Sequence: Confusion of Patterns Other Than Speech and Music," *Science*, Vol. 184, 1969, pp. 586–587.
67. I. J. Hirsh, "Auditory Perception of Temporal Order," *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 759–767.
68. R. M. Warren, *Auditory Perception: A New Synthesis*, Pergammon, New York, 1982.

69. W. J. Kelly and C. S. Watson, "Stimulus-Based Limitations on the Discrimination between Different Temporal Orders of Tones," *J. Acoust. Soc. Am.*, Vol. 79, 1986, pp. 1934–1938.
70. M. R. Jones, "The Patterning of Time and Its Effects on Perceiving," *Ann. N.Y. Acad. Sci.*, Vol. 423, 1984, pp. 158–167.
71. M. R. Jones, M. Boltz, and G. Kidd, "Controlled Attending as a Function of Melodic and Temporal Context," *Percept. Psychophys.*, Vol. 32, 1982, pp. 211–218.
72. M. R. Jones, G. R. Kidd, and R. Wetzel, "Evidence for Rhythmic Attention," *J. Exper. Psychol.: Human Percept. Perform.*, Vol. 7, 1981, pp. 1059–1073.
73. R. D. Sorkin, "Temporal Factors in the Discrimination of Tonal Sequences," *J. Acoust. Soc. Am.*, Vol. 82, 1987, pp. 1218–1226.
74. G. R. Kidd, M. Boltz, and M. R. Jones, "Some Effects of Rhythmic Context on Melody Recognition," *Am. J. Psychol.*, Vol. 97, 1984, pp. 153–173.
75. R. D. Sorkin, G. J. Boggs, and S. L. Brady, "Discrimination of Temporal Jitter in Patterend Sequences of Tones," *J. Exper. Psychol.: Human Percept. Perform.*, Vol. 8, 1982, pp. 46–57.
76. K. W. Berger, "Some Factors in the Recognition of Timbre," *J. Acoust. Soc. Am.*, Vol. 36, 1964, pp. 1888–1891.
77. L. Wedin and G. Goude, "Dimension Analysis of the Perception of Instrumental Timbre," *Scand. J. Psychol.*, Vol. 13, 1972, pp. 228–240.
78. P. Bricker and S. Pruzansky, "Speaker Recognition," in N. Lass (Ed.), *Contemporary Issues in Experimental Phonetics*, Academic, New York, 1976, pp. 295–326.
79. D. Van Lancker, J. Kreiman, and K. Emmorey, "Familiar Voice Recognition: Patterns and Parameters. Part I: Recognition of Backward Voices," *J. Phonet.*, Vol. 13, 1985, pp. 19–38.
80. J. Jenkins, W. Strange, and T. Edman, "Identification of Vowels in "Vowelless" Syllables," *Percept. Psychophys.*, Vol. 341, 1983, pp. 441–450.
81. W. Strange, J. J. Jenkins, and T. L. Johnson, "Dynamic Specification of Coarticulated Vowels," *J. Acoust. Soc. Am.*, Vol. 74, 1983, pp. 695–705.
82. D. Kewley-Port, "Time-Varying Features as Correlates of Place of Articulation in Stop Consonants," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 322–335.
83. W. W. Gaver, Auditory Icons: Using sound in computer interfaces. *Human Computer Interaction*, Vol. 2, 1986, pp. 167–177.
84. R. H. Lyon, "Engineering for Sound Quality," in J. Cuschieri, S. Giegg, and D. Yeager (Eds.), *Proceedings of Noise Control 94*, Ft. Lauderdale FL, May 1–4, 1994, Noise Control Fdn., New York, 1994, pp. 3–8.
85. W. H. Warren and R. R. Verbrugge, "Auditory Perception of Breaking and Bouncing Events," *J. Exper. Psychol.: Human Percept. Perform.*, Vol. 10, 1984, pp. 704–712.
86. B. H. Repp, "The Sound of Two Hands Clapping: An Exploratory Study," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 1100–1109.
87. X. Li, R. J. Logan, and R. E. Pastore, "The Perception of Complex, Naturally Occurring, Nonspeech Auditory Stimuli: Walking Perception," *J. Acoust. Soc. Am.*, Vol. 87, Suppl. 1, 1990, p. S24 (abst.).
88. D. Freed, "Auditory Correlates of Perceived Mallet Hardness for a Set of Recorded Percussive Sound Events," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 311–322.
89. N. J. Van Derveer, "Acoustic Information for Event Perception," paper presented at the celebration in honor of Eleanor J. Gibson, Cornell University, Ithaca, NY, 1979.
90. W. W. Gaver, "What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception," *Ecol. Psychol.*, Vol. 5, 1993, pp. 1–29.
91. W. W. Gaver, "How Do We Hear in the World?: Explorations in Ecological Acoustics," *Ecol. Psychol.*, Vol. 5, 1993, pp. 285–313.
92. J. J. Jenkins, "Acoustic Information for Objects, Places, and Events," in W. H. Warren and R. E. Shaw (Eds.), *Persistence and Change*, Lawrence Erlbaum, Hillsdale, NJ, 1985, pp. 115–138.
93. C. S. Watson, "Time Course of Auditory Perceptual Learning," *Ann. Otol. Rhinol. Laryngol.*, Vol. 74, 1980, pp. 96–102.
94. C. S. Watson, "Auditory Perceptual Learning and the Cochlear Implant," *Am. J. Otol.*, Vol. 12, Suppl., 1991, pp. 73–79.
95. J. Zwislocki, F. Maire, A. S. Feldman, and A. Rubin, "On the Effect of Practice and Motivation on the Threshold of Audibility," *J. Acoust. Soc. Am.*, Vol. 30, 1958, pp. 254–262.
96. R. A. Campbell and A. M. Small, Jr., "Effect of Practice and Feedback on Frequency Discrimination," *J. Acoust. Soc. Am.*, Vol. 35, 1963, pp. 1511–1514.
97. M. L. Meyer, "Is the Memory of Absolute Pitch Capable of Development by Training?" *Psycholog. Rev.*, Vol. 6, 1899, pp. 514–516.
98. R. F. Gundy, "Auditory Detection of an Unspecified Signal," *J. Acoust. Soc. Am.*, Vol. 33, 1961, pp. 1008–1012.
99. M. Loeb and D. H. Holding, "Backward Interference by Tones or Noise in Pitch Discrimination," *Percept. Psychophys.*, Vol. 18, 1975, pp. 205–208.
100. W. P. Tanner, Jr. and C. L. Rivette, "Learning in Psychophysical Experiments," *J. Acoust. Soc. Am.*, Vol. 35, 1963, p. 1896 (abst.).
101. C. S. Watson, J. R. Franks, and D. C. Hood, "Detection of Tones in the Absence of External Masking Noise," *J. Acoust. Soc. Am.*, Vol. 52, 1972, pp. 633–643.
102. C. S. Watson and W. J. Kelly, "Discrimination of Tonal Patterns with Several Levels of Stimulus Uncertainty," *J. Acoust. Soc. Am.*, Vol. 64, 1978, p. S39 (abst.).
103. E. B. Hartman, "The Influence of Practice and Pitch-Distance between Tones on the Absolute Identification of Pitch." *Am. J. Psychol.*, Vol. 67, 1954, pp. 1–14.
104. C. S. Watson and W. J. Kelly, "Informational Masking in

Word-Length Tonal Patterns," *J. Acoust. Soc. Am.*, Vol. 64, 1978, p. S39 (abst.).

105. R. M. Warren, "Auditory Temporal Discrimination by Trained Listeners," *Cogn. Psychol.*, Vol. 6, 1974, pp. 237–256.
106. W. L. Bryan and N. Harter, "Studies in the Physiology and Psychology of the Telegraphic Language," *Psychol. Rev.*, Vol. 6, 1899, pp. 345–375.
107. M. J. Leek and C. S. Watson, "Auditory Perceptual Learning of Tonal Patterns," *Percept. Psychophys.*, Vol. 43, 1988, pp. 389–394.
108. M. R. Leek and C. S. Watson, "Learning to Detect Auditory Pattern Components," *J. Acoust. Soc. Am.*, Vol. 76, 1984, pp. 1037–1044.
109. B. H. Deatherage, "Auditory and Other Sensory Forms of Information Presentation," in H. P. Van Cott and R. G. Kinkade (Eds.) *Human Engineering Guide to Equipment Design*, U.S. Dept. of Defense, Army Res. Inst., Washington, DC, 1972, pp. 123–160.
110. R. D. Patterson, "Guidelines for Auditory Warning Systems on Civil Aircraft," Civil Aviation Authority, Paper 82-017, London, 1982.
111. R. D. Sorkin, "Design of Auditory and Tactile Displays," in G. Salvendy (Ed.), *Handbook of Human Factors*, Wiley, New York, 1987, pp. 449–576.
112. K. R. Boff and J. E. Lincoln, *Engineering Data Compendium: Human Perception and Performance*, AAMRL, Wright-Patterson AFB, OH, 1988.
113. R. D. Patterson and R. Milroy, "Auditory Warnings on the Boeing 747, Civil Aviation Authority, Contract No. 7D/S/0260/1, Stage II, Final Report, 1982.
114. R. D. Patterson, "Auditory Warning Sounds in the Work Environment," *Philos. Trans. Roy. Soc. Lond.*, Vol. B327, 1990, pp. 485–492.
115. Accredited Standards Committee S3, Bioacoustics, "American National Standard Audible Emergency Evacuation Signal," Acoustical Society of America, New York, 1991.

122

ADAPTATION IN THE AUDITORY SYSTEM

PAUL J. ABBAS

1 INTRODUCTION

The term *adaptation* refers to a decrease in responsiveness that occurs during and following the presentation of a stimulus. Both the psychophysical and physiological response of the auditory system can change during the presentation of an acoustic stimulus that is constant over time. Similarly, after the offset of a stimulus, the auditory system may require a certain period of time before recovery from stimulation is complete. In psychophysical experiments, adaptation is evidenced by changes in the perceived loudness of a stimulus or in the changes of the threshold detection level of a stimulus over time. Physiological measures demonstrate changes over time in the discharge rate of single neurons and also in the amplitude of evoked potentials from both the auditory nerve and higher levels of the auditory nervous system. Possible sources of adaptation within the auditory system and models that relate the physiological measures to changes in perception will be presented. Some authors have made a distinction between adaptation and fatigue; fatigue involves stimulation above "normal" physiological levels and requires a relatively long poststimulus recovery period. In this chapter we will focus on adaptation rather than fatigue.

2 PSYCHOPHYSICAL MEASURES

Psychophysical measures of adaptation can be either perstimulatory (during the presentation of the adapting stimulus) or poststimulatory (following the offset of the adapting stimulus). Perstimulatory data have primarily been obtained using one of two methods: direct assessment of the loudness of the adapting stimulus over its time course and measurement of psychophysical threshold for a "probe" stimulus presented simultaneously with the adapting stimulus. Poststimulatory measures typically require measurement of detection threshold for a probe presented at various times following offset of the adapter in order to assess the time course of recovery from adaptation. The sections below outline the results of experiments using each of these measurement paradigms.

2.1 Loudness Adaptation

The loudness of an acoustic stimulus is a perceptual variable that depends strongly on the sound pressure level (SPL) of the stimulus. Loudness can be measured directly by requiring the subjects to assign numbers to a given stimulus corresponding to their perception of the loudness of that signal, a procedure known as magnitude estimation. A second, less direct method of measuring loudness is to require the subjects to adjust the level of a probe to match the loudness of the stimulus.

The extent of loudness adaptation that occurs is dependent on the specific stimuli and the method used to assess adaptation. Scharf[1] has summarized a number of experiments in which he measured loudness adaptation using the technique of *successive magnitude estimation*. This technique is probably the most straightforward way of assessing the loudness of a stimulus. In those experiments, subjects are required to assign a number corresponding to the perceived loudness of a stimulus at successive intervals during the continuous presentation of the stimulus. Using this method, loudness estimates of a pure-tone stimulus tended to decrease over the course of several minutes for a low-level probe. Loudness esti-

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

mates could be as low as 10% of the original estimates after a 7 dB sensation level (SL) stimulus was on for 30 s; there was little or no change in the loudness estimates for stimulus levels above 30 dB SL. The spectral properties of the stimulus can affect adaptation; high-frequency pure tones generally show more adaptation than low-frequency tones. Little adaptation was found when broadband noise or filtered noise was used. Less adaptation was found for adapter composed of two tones spaced close in frequency than when the adapter consisted of two widely spaced tones.

Other methods of assessing loudness adaptation use comparisons with other stimuli to assess loudness. The subject's task is to adjust the level of the probe stimulus until it is equal in loudness to that of the adapter. Investigators who have used a *simultaneous dichotic loudness balance*[2] (the probe and adapter are presented simultaneously to opposite ears) have observed substantial amounts of adaptation. Investigators using a *delayed dichotic loudness balance* procedure in which the probe tone is presented after the adapting tone is turned off have observed little or no adaptation; that is, loudness matches for short-duration adapters are similar to those obtained when long-duration adapters are used.[3] A second technique is to use a probe that is different in frequency from the adapter. This technique is called *simultaneous heterophonic loudness balance*. Studies using this technique with the probe and the adapter in opposite ears have generally shown little or no evidence of adaptation.[4] Studies using similar techniques in which the stimuli are presented monaurally have shown as much as 10–20 dB of adaptation.[5] Using scaling procedures, Canevet et al.[6] have demonstrated that significant adaptation can be induced by intermittent stimulation (in either ear), which is typically an element of loudness-matching paradigms.

In general, measures of loudness adaptation depend strongly on the method used. Most methods demonstrate some degree of adaptation taking place over a time course that is on the order of minutes. For monaural stimulus presentations, adaptation is generally limited to low-level stimuli. When loudness-matching methods are used, the results are particularly dependent on the stimulus paradigm.

2.2 Simultaneous Masking

Perstimulatory adaptation can also be assessed by measuring the threshold of a probe that is presented during the adapting stimulus, as illustrated in Fig. 1*a*. Since the adapter increases the threshold for the probe, in much of the literature it is termed the masker. The probe stimulus typically has a relatively short duration, and its onset relative to that of the adapter/masker (Δt) is varied. The detection threshold of the probe is higher when the signal is presented near adapter onset ($\Delta t = 2, \cdots, 20$ ms) than when the Δt is on the order of 200–300 ms.[7] The term "overshoot" has been used to describe this phenomenon. Detection threshold for the probe measured as a function of Δt is plotted in Fig. 2. Functions such as that shown in Fig. 2 tend to reach an asymptote with 100 ms after adapter onset.[8]

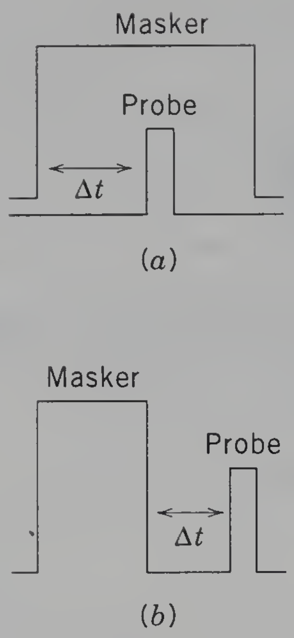


Fig. 1 Schematic diagrams showing stimuli used in listening experiments; Δt time between adapter onset and probe onset: envelope of stimulus presentation in (*a*) masking studies that measure the overshoot effect and (*b* studies of forward masking.

The amount of overshoot, or the difference in the amount of masking that occurs for a probe presented at the onset and that obtained when the probe is presented during the steady-state part of the masker, can be as high as 25 dB in experiments in which both the adapter and probe are pure tones.[9] At least part of the effect may be related to the spectral splatter that accompanies the use of a fast rise-time envelope. More recent experiments in which the gating was better controlled to reduce such frequency splatter have yielded values of overshoot that are on the order of 15 dB.[10] Similar amounts of overshoot have been observed with tonal probes and noise maskers.[11] With tonal stimuli, the overshoot effect depends on the relative frequency of the masker and the probe. When both the masker and probe are equal in frequency, there is little overshoot. For probe frequencies below the adapter frequency, the amount of overshoot is greater and the steady-state threshold is reached when Δt is about 100 ms.[8] With noise maskers, relatively large overshoot effects are observed only for wide-band maskers.[12]

The amount of overshoot is highly dependent on

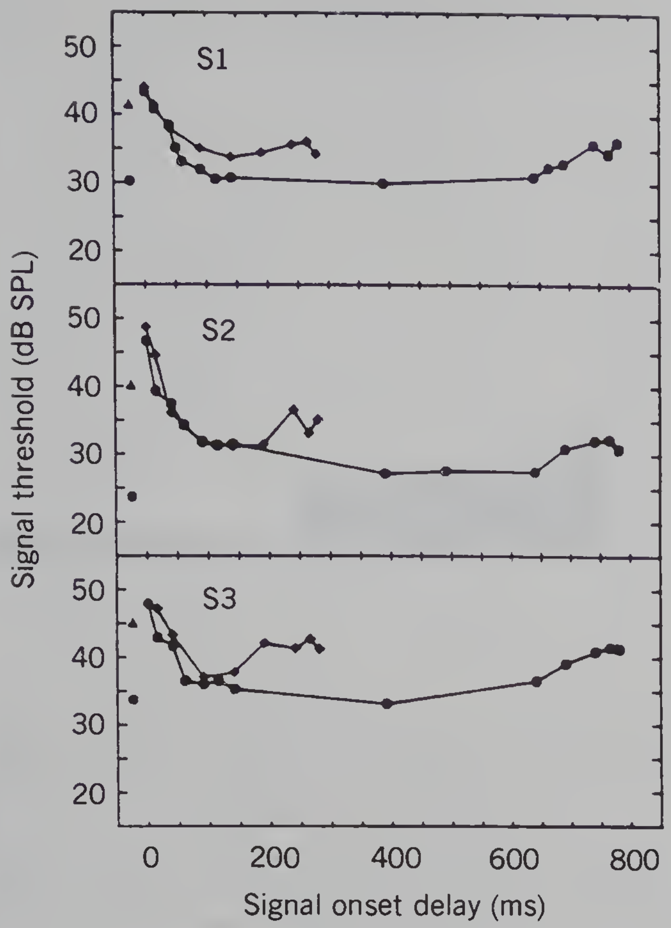


Fig. 2 Threshold as a function of signal onset delay (Δt) plotted for three subjects. The adapter was a 1250-Hz tone at 80 dB SPL. Its duration was either 300 ms (♦) or 800 ms (●). The probe was a 20-ms, 1-kHz tone. Thresholds for a 50-ms adapter (▲) and for a continuous adapter (●) are also shown. Reprinted with permission from Ref. 8.

the level of the adapter.[13] When a continuous noise masker is used or the probe tone is presented 200 ms or more after the masker onset, the detection threshold for the probe increases approximately linearly with masker level. When the probe is presented near the onset of a gated masker, the growth of the probe detection level is greater than linear at low to moderate levels but less than linear at high levels. Thus, at very low levels and at high levels for most subjects, there is little overshoot. For moderate noise spectrum levels of 10–40 dB SPL, the amount of overshoot reaches a maximum of about 15 dB.[13]

Detection tasks in which the probe and masker are spectrally complex have demonstrated adaptation effects that are somewhat different from the overshoot effects described above. Green et al.[14] observed that the prominence of a particular frequency in a multitone complex can be enhanced by prior stimultion with a similar complex with that frequency component removed. Viemeister[15] quantified this effect with a number of different stimuli (either harmonic complexes with one frequency removed or band-reject noise). Probe stimuli were either pure tones at the missing frequency or bandpass noise. He found that probe threshold measured using continuous maskers were on the order of 10 dB lower than those obtained using gated maskers. By presenting the masking stimulus for different durations before the probe, the time course of threshold decrease was shown to be approximately 400 ms, somewhat longer than the time course observed in overshoot experiments.

McFadden and Wright[16] measured the detectability of a bandpass noise probe in the presence of a masker consisting of several nonoverlapping flanking bands. They controlled the time envelope of the masker and probe stimuli and presented conditions in which the masker and probe envelopes were either correlated or uncorrelated. The difference in detectability in these conditions is referred to as *comodulation detection difference*.[17] When the masking bands are present prior to the onset of the probe band, the detection threshold tends to decline as Δt is increased. The effect is subject dependent, but in those subjects who show a reduction in threshold, the decrease takes place over a range of 300–400 ms. Threshold differences across Δt were on the order of 20 dB between 0 and 700 ms. In both of these masking paradigms, using specifically complex stimuli, the time course of threshold change is longer than typically observed with overshoot effects.

2.3 Poststimulatory Threshold Shift

The most common measurements of the poststimulatory effects of adaptation are made using detection threshold paradigms. For low-level, short-duration adapters, the recovery of threshold to normal values is typically less than a second and the threshold elevation is termed forward masking. A typical stimulus paradigm illustrated in Fig. 1*b* consists of an adapter/masker followed by a silent interval (Δt) and then a probe stimulus. This section summarizes the effects of Δt, adapter level, adapter duration, and adapter and probe frequency.

Threshold for detection of a probe (in decibels) plotted as a function of Δt is usually used to describe recovery from adaptation in psychophysical experiments. In forward masking, the recovery function is monotonic; the function is approximately exponential (linear when plotted on scale of decibels threshold shift vs. logarithm of Δt).[18] The time required for recovery from adaptation is dependent on the duration and level of the adapter. For adapters less than 200 ms in duration and less than 80 dB SPL, recovery is generally complete within several hundred milliseconds. Duifhuis[19] proposed that recovery functions might better be described by the sum of two exponential functions, one with a relatively short time constant (<10 ms) and one with a relatively long time constant (approximately 75 ms). Widin and Viemeister[20] measured recovery functions for different levels of adapters and observed a different slope of recovery (decibels vs. log Δt plot), indicating a different time constant of exponential recovery for each adapter level.

As the level of the adapter increases, the threshold of a probe measured at a specific Δt also generally increases. In forward masking, the probe threshold grows with a slope of less than 1 relative to the adapter; that is, probe threshold increases more slowly than adapter level. The slope of this masking function tends to decrease as the frequency of the adapter is increased relative to that of the probe. The slope of the masking function also decreases as the value of Δt is increased.[20]

Increasing the duration of an adapter in a forward-masking paradigm can increase the amount of masking.[21] The most effective adapter frequency is at or near the frequency of the probe.[22]

The characteristics of the recovery process and the dependence of adaptation on the parameters of the adapting stimulus are highly dependent on the range of the specific paramgers. The effects described above are simplified to the extent that there are interactions among the individual variables discussed. For instance, the time course of recovery and the effect of adapter duration may depend on the frequency of the adapter.[23]

3 PHYSIOLOGICAL MEASURES

This section outlines physiological studies that parallel the psychophysical studies summarized in the previous sections. The following subsections are organized into the type of physiological measures that have been used to assess adaptation. Many physiological adaptation studies have focused on the responses of single nerve fibers of the auditory nerve and on far-field measures of the whole-nerve activity. The final section deals with some of the mechanisms that may be involved in producing adaptation. Little adaptation is observed in the physiological responses peripheral to the auditory nerve; intracellular responses from hair cells in the cochlea show little adaptation in response to low-level, short-duration stimuli.[24]

3.1 Single Nerve Fiber Responses

Perstimulatory Measures The responses of neurons of the auditory nerve can show substantial adaptation. A poststimulus time (PST) histogram displays the average nerve discharge as a function of time during stimulus presentation.[25,26] An example is shown in Fig. 3. Histograms obtained in response to tones or noise bursts show a high probability of generating an action potential at the onset of the stimulus. The probability of firing quickly decreases to a steady-state response rate (actually slowly decreasing over the course of minutes). After the onset of the stimulus, there is a short period of reduced spontaneous activity during which the nerve fiber recovers to its normal spontaneous rate. The rise time of the tone burst affects the relative amplitude of the onset and steady-state portions of the response. Faster rise time results in a greater onset effect.[27] The details of perstimulatory adaptation have been characterized as a three-segment function: a very fast period of adaptation that occurs during the first few milliseconds after stimulus onset, a slower decrease in firing rate that occurs over the next 50–100 ms, and a steady-state portion of the response histogram. The PST histogram can be fitted by two decreasing exponential functions with different time constants, one on the order of 10 ms and the second on the order of 100 ms.[28,29] At stimulus levels near fiber threshold, the histograms show little onset activity.[26] At stimulus levels within the fiber's dynamic range, the three segments of the function become noticeable. As stimulus level is increased further, the second and third segments reach saturation while the initial peak in the response continues to grow.[30]

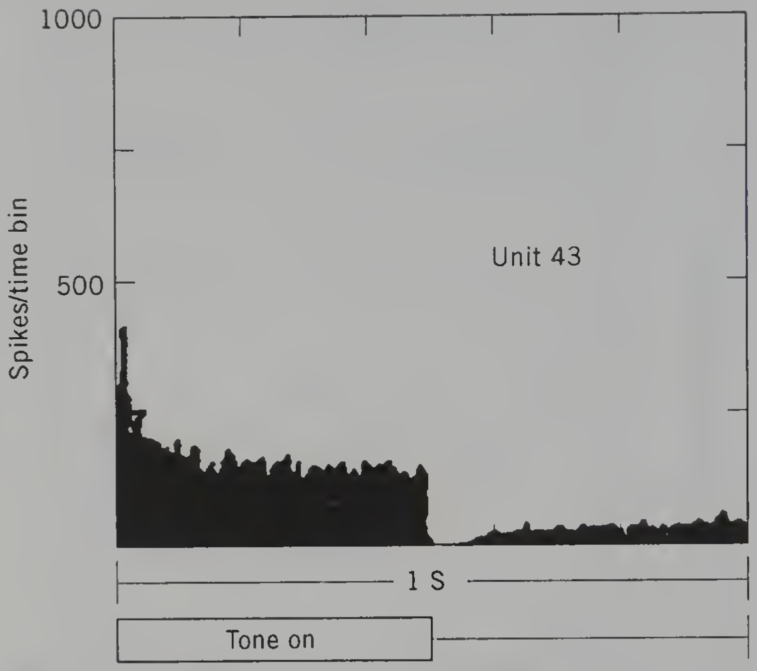


Fig. 3 A PST histogram and typical of a high spontaneous rate auditory nerve fiber. Number of action potentials is plotted as a function of time relative to stimulus onset. The stimulus envelope is illustrated below. Auditory nerve fibers with low spontaneous rate typically show less decrease over time.[39] (From Kiang et al., *MIT Monographs*, No. 35, 1965. Copyright 1965 MIT Press. Reprinted with permission.)

Smith and Zwislocki[31] observed the effects of perstimulatory adaptation properties in auditory nerve fiber responses. Their stimulus was a tone burst, termed the

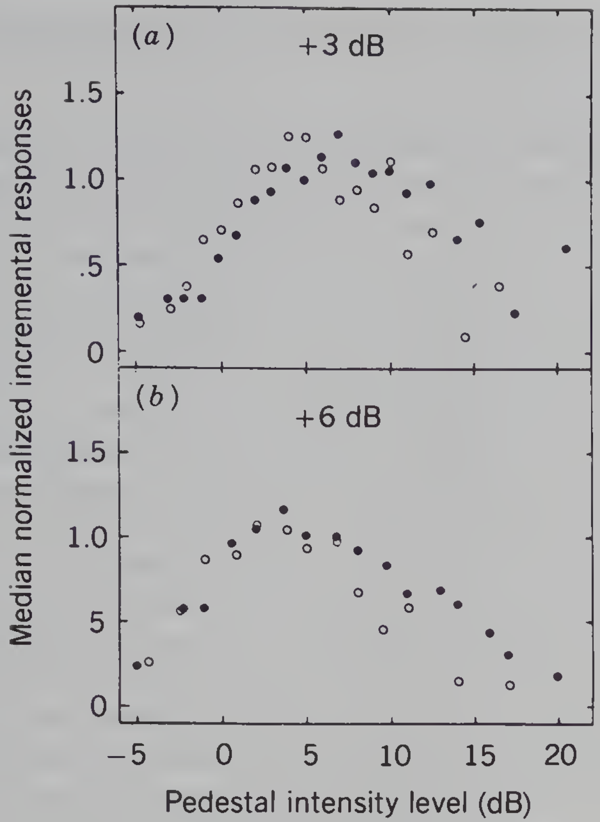


Fig. 4 Values of the measured increase in response to the level increment as a function of pedestal (adapter) level. Responses for an increment at adapter onset (open symbols) and one during adapter steady state (filled symbols) are shown for increment values of 3 and 6 dB. These are median results obtained by normalizing data from 10 individual neurons and the computing the median values. Pedestal level (dB) is referenced to threshold for each individual neuron. (Reprinted with permission from Ref. 31.)

pedestal, in which the level was increased at a certain time after stimulus onset. They measured the increase in discharge rate in response to each increment in sound pressure level at adapter onset and during the steady-state portion of the adapter (Δt = 150 ms). Results from their experiments are shown in Fig. 4. The increment in discharge rate is plotted as a function of pedestal sound pressure level for two values of level increments (3 and 6 dB). The increased discharge rate that follows each increment in level is dependent on pedestal level but is generally the same at the onset of tone burst as during the steady-state portion of the burst (open and closed symbols). For each level, response to the adapter (or pedestal) 150 ms after onset has decreased considerably (as in Fig. 4), but the increment in response is not diminished.

If one considers the increment in level as a probe stimulus and the adapter as a masking stimulus in a psychophysical task, the increment in response is the same for the onset and steady-state portions of the adapter, but the background noise or response to the masker has adapted. Thus, the neural *signal-to-noise ratio* is better later than earlier in the pedestal, an effect qualitatively similar for psychophysical overshoot. A second important feature of the data in Fig. 4 is that the incremental response is small for low adapter levels, increases as adapter level is increased, and then decreases at higher adapter levels. That is, the fibers are sensitive to the increase in response only over a fairly narrow range of adapter levels, independent of the timing of the increment.

Poststimulatory Measures Recovery from adaptation also can occur over very different time scales, depending on the level and duration of the adapter. The recovery from adaptation is generally measured using a short probe stimulus in an experimental paradigm similar to that shown in Fig. 1*b*. In general, the discharge rate to the probe is decreased relative to the unadapted case. If the adapting stimulus is short (approximately 1 s or less), then recovery is quick and is usually complete within several hundred milliseconds. The recovery function is approximately exponential in form; the magnitude of the decrease becomes greater as either discharge rate to the adapter or the duration of the adapter is increased.[32,33] The spontaneous discharge rate after stimulus offset (see Fig. 3) shows a recovery similar in time course to the probe discharge rate.[32] Recovery time constants vary across animal species, typically from 25 to 200 ms.[32,33] The time constant of recovery increases with both increasing adapter level and adapter duration.[32] Nevertheless, the time constant for recovery from adaptation is consistently greater than the time constant of functions describing the perstimulatory decrease in response with time. That is, adaptation is faster than recovery.[32,33]

Figure 5 illustrates the effects of changing stimulus level. As adapter level is increased, the increased response to the adapter results in a decrease in discharge rate to the probe (test tone). As the level of the adapting stimulus is increased to the point of response saturation, the probe (test tone) response similarly "saturates"; that is no further decreases in discharge rate are observed. Thus, for a particular Δt, the decrement in discharge rate to the probe is simply proportional to the adapter response. Furthermore, as other adapter stimulus parameters are varied, the decrement in response to the probe simply reflects the changes in response to the adapter.[32–34]

The response of neurons with different rates of spontaneous activity differ in a number of ways. Low-spontaneous-rate neurons have a higher threshold of response than high-spontaneous-rate neurons.[35] Adaptation properties also differ in that PST histograms of low-spontaneous-rate neurons show little perstimulatory adaptation compared to those of high-spontaneous-rate neurons.[26] Relkin and Doucet[36] have observed that post-

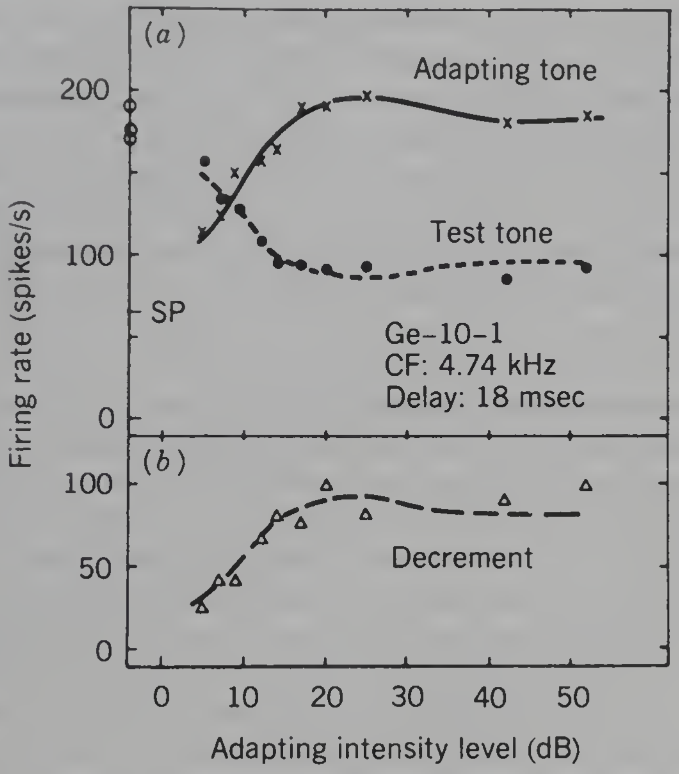


Fig. 5 (*a*) Average discharge rate to an adapting tone is plotted as a function of adapting tone level (x). A probe is presented after offset of the adapter (Δt = 18 ms) at a constant level. Discharge rate to the probe (test tone) is plotted on the same scale (●). (*b*) Decrement in discharge rate to the probe (response to the probe alone minus response to the probe with a preceding adapter) is plotted as a function of adapter level (△). (Reprinted with permission from Ref. 33.)

stimulatory recovery in low-spontaneous-rate neurons was slower (sometimes incomplete by 300 ms) than in high-spontaneous-rate neurons. In some ways we may view the auditory nerve as two subpopulations of neurons, each with a different threshold sensitivity and adaptation properties.

3.2 Far-Field Measures

Far-field measures of brain and nerve activity in the auditory system include the compound action potential (CAP) and the auditory brainstem response (ABR). The CAP represents activity from the population of auditory nerve fibers, and the ABR includes responses from the auditory nerve and nuclei of the lower auditory pathway.[37,38] Far-field measures can be useful in viewing an overall response pattern of a population of neurons but also are particularly useful in providing a link between the physiological responses obtained in animals and the types of responses that can be recorded in humans. Measurement of both the CAP or auditory brainstem response is feasible using noninvasive techniques. In recording the CAP, responses to several hundred identical stimuli are typically averagesd in time to enhance the response potential relative to subject- and equipment-generated electrical noise.

Perstimulatory Measures The CAP is typically an onset response where synchronously firing action potentials at stimulus onset produce this measurable far-field response. Consequently, these measures are not typically used for assessing perstimulatory changes in response, but several variations of these methods in experimental animals have proved useful for this purpose. Chimento and Schreiner[39] have described a technique to measure whole nerve activity with a bipolar electrode placed directly in the auditory nerve. They termed the measurement the *auditory nerve neurophonic potential*, which is the spatially summed responses of phase-locked auditory nerve fibers. They have used the technique to measure perstimulatory adaptation effects and have fitted the adaptation function with a two-exponential model similar to those used to describe single-fiber responses.[28,29] Relkin and Doucet[40] have also measured perstimulatory effects using an electrode that applies suction directly to the nerve and presumably blocks the action potentials from propagating past it. The contribution of each action potential to the recorded potential is then monophasic, resulting in DC offset during periods of sustained activity. Both methods yield perstimulatory adaptation functions similar to single-nerve fiber recordings.

Poststimulatory Measures The CAP is well-suited for assessment of poststimulatory adaptation effects.[41,42] With brief tonal adapters, the CAP response to tone-burst probes shows a brief decrease in amplitude that is similar to the changes in discharge rate of auditory nerve fibers under similar stimulus conditions.[42] Relkin and Smith[43] have used a two-alternative, forced-choice paradigm to measure threshold of detection of the CAP elicited by a probe stimulus. They measured growth-of-masking functions (probe threshold level vs. adapter level) and observed functions similar to human psychophysical data in that the slope decreased with increasing Δt (see Section 2.3).

Frequency effects are also similar to psychophysical data. When the adapter frequency is equal to that of the probe, the greatest decrement in probe response is observed; frequencies above or below the probe frequency show decrements consistent with the tuning properties of auditory nerve fibers.[44] Similar effects have also been observed for the auditory brainstem response.[45]

3.3 Sources of Adaptation

The specific sources of adaptation and fatigue in the auditory system have not yet been established, but several sources are thought to contribute to them. Since short-term adaptation is not observed in hair cell electrical potentials, the sources of adaptation are considered to be more central in the auditory pathway. The receptor potential of the hair cell controls the release of neurotransmitter and consequently affects action potential generation in auditory nerve fibers. Short-term adaptation as measured in these neurons is most likely related to the physiological properties of the synapse. Recordings of postsynaptic potentials in goldfish have demonstrated adaptation of slow postsynaptic potentials in response to acoustic stimulation.[46] Adaptation-like effects have been observed in single-nerve fiber responses to electrical stimulation of the cochlea. Javel et al.[47] used 100-ms trains of pulses and at certain stimulus levels observed a decrease in the PST histogram over time. van den Honert and Stypulkowski[48] used sinusoidal electrical stimulation and observed a decrease in the discharge rate over a period of several seconds. In both cases, the adaptation effects were not normal vis-à-vis acoustic stimulation, but the results implicate mechanisms of adaptation in addition to synaptic transmitter depletion.

A number of models have been developed in which the physiological response properties can be quantitively described by simulations of specific synaptic and neuronal properties. The simplest model of the hair cell–neuron synapse is a single reservoir in which the depletion of neural transmitter results in lower synaptic output and consequently adaptation of discharge rate.[49] Limitations of these models have led others[30,50] to develop multireservoir models in which the release of transmitter at specific sites depends on stimulus level. As level is increased, transmitter is released at more sites. Meddis[51] described a single-reservoir model but incorporated a quick uptake of released transmitter for reuse by the presynaptic membrane. The model also incorporates some loss of the transmitter in the cleft, which contributes to its adaptation properties. These model assumptions fit many of the known adaptation properties without the complexity of a multireservoir model.[52] Other models have taken into account the dynamic properties of receptor sites in the postsynaptic membrane.[53] Finally, the refractory properties of the neuron itself can affect adaptation observed in the PST histogram.[29,53]

4 RELATIONSHIPS BETWEEN PHYSIOLOGICAL AND PSYCHOPHYSICAL MEASURES

As can be seen from the previous sections, there are many similarities between psychophysical and physiological measures. For instance, the discharge rate of some neurons decreases over the time of stimulus presentation and the perceived loudness of a stimulus can also decrease. Since the response to the adapting stimulus is greater at stimulus onset and the incremented (probe) response does not change, signal-to-noise measures are consistent with more effective masking at onset than at later times. Finally, poststimulatory changes in discharge rate show a time course of recovery similar to psychophysical forward-masking functions. Threshold measures of single-fiber discharge rate and CAP amplitude using two-alternative, forced-choice paradigms show changes in slope of the masking function with increasing Δt that are similar to those observed psychophysically. Both threshold shift and discharge rate of neurons tend to reach asymptotic values as the duration of the adapter increases up to approximately 1 s. The decrease in discharge rate and in CAP amplitude with increasing adapter level is dependent on the adapter frequency. Rate of growth of psychophysical masking functions show similar changes with adapter frequency. Changes in loudness estimates are also dependent on the relative frequency of the adapter and probe tones. More quantitative inspection of the data shows further similarities but also some important differences between physiological and psychophysical data.

Smith and Zwislocki[31] (Section 3.1) made specific comparisons of their physiological data to psychophysical masking data. They assumed that the signal-to-noise ratio in discharge rate determines the detectability of the signal (rate to the adapter pedestal equals the noise, rate to the increment equals the signal). They noted that a single fiber gives useful information over only a small range of adapter intensities and concluded that fibers with characteristic frequencies different from the stimulus would provide such information. On the basis of discharge rate adaptation, they calculated that detectability of a probe would decrease 3–5 dB from stimulus onset to the steady-state portion of the signal. Cases where the adapter and probe are equal show similarly small psychophysical effects. Large psychophysical effects observed with adapter frequency greater than the probe or where broadband maskers are used cannot be accounted for by this model.

It is particularly important to take into account the changes both in spontaneous rate and in response that are measured in poststimulatory adaptation paradigms. The decrease in response is accompanied by a similar decrease in the spontaneous rate of high-spontaneous-rate fibers so that a proportional threshold change might not follow. Relkin and Turner[54] addressed this issue directly by collecting the physiological data using a two-interval, forced-choice detection paradigm. Probe level was adaptively adjusted to threshold using a criterion of change in discharge rate. The adapting signal was effective in reducing the threshold to the probe tone, but in general, the changes in threshold were less robust than those observed psychophysically. Similar experiments using the CAP with the same stimulus paradigm have shown growth of masking similar to psychophysical performance[43] (Section 3.2).

Other studies have developed more detailed models to relate physiological and psychophysical data. One fairly straightforward way of comparing the data is to model the central nervous system as an ideal detector and to ask what are the limitations on performance based on the response properties of single nerve fibers.[55] Young and Sachs[56] developed such a model to compare the recovery data obtained in psychophysical and physiological experiments. They assumed that discharge rate of neurons was the criterion measure and they compared rates with and without the probe stimulus present to calculate thresholds. The model reconciled some differences between the physiological and psychophysical data, but changes in the psychophysical time constant with adapter level could not be simulated. They suggested that one explanation for the differences was central neural adaptation.

The physiological data currently available are primarily from the auditory periphery; certainly the changes seen at the periphery will influence subsequent processing and ultimately perception. Nevertheless, there is increasing evidence that the central processing and ultimately perception. Nevertheless, there is increasing evidence that the central processing of these signals also contributes to the effects of adaptation observed psychophysically. McFadden and Champlin[57,58] have outlined a number of limitations to the peripheral physiological model based on the nerve responses as an explanation for psychophysical overshoot effects. The amount of overshoot is different when predicted from physiological rather than psychophysical adaptation (see above) and the psychophysical functions show faster adaptation than PST histograms. Another difference is that the physiological onset effects tend to increase as a function of adapter level, whereas the psychophysical overshoot effects increase, diminish, and finally vanish with increasing stimulus levels. One explanation for this difference is that detection may be based on the tapping of different populations of nerve fibers under different stimulus conditions.[13,50] High-spontaneous-rate neurons are more sensitive fibers and show adaptation and therefore also overshoot, particularly at high stimulus levels. At higher stimulus levels detection may be based on the responses of less sensitive, low-spontaneous-rate fibers with response rates yet unsaturated. Those fibers show little or no adaptation effects when small ISIs are used.[26] As noted in Section 3.1, however, the poststimulatory recovery of fibers with low-spontaneous-rate fibers is slower than those with high spontaneous rate, which accounts for these differences in perstimulatory measurements. Zeng et al.[59] have argued that these two populations may be responsible for suprathreshold responses after the adapter (in particular, intensity discrimination) but are not likely to affect detection tasks in quiet.

A final limitation of the peripheral physiological overshoot model is related to frequency effects. The auditory system can recover from overshoot effects when a temporal gap is placed in the adapting stimulus.[11] McFadden[60] extended those experiments using combinations of noise bands centered on and flanking the probe stimulus. Each of the bands was gated differentially. The recovery of overshoot depended on the gating of the noise bands away from the probe rather than the band at the same frequency as the probe. These results suggest that processing in the central nervous system may be an important determinant in overshoot under such stimulus conditions.

The above observations are relevant to the relationship between overshoot effects and physiological adaptation. Several observations related to forward masking also implicate central effects underlying psychophysical performance. As with simultaneous masking, neither the time course nor the amount of physiological adaptation is consistent with auditory nerve fiber response data. Shannon[61] has recently reported forward-masking data on patients who have their auditory nerve directly stimulated through an implanted electrode array in the cochlea (cochlear implant). Since these subjects likely have few or no functioning hair cells, they lack at least a major part of the adaptation mechanism that affects the physiological responses of the auditory nerve. The perceptual growth functions and dynamic ranges of these subjects using electrical stimulation is very different from that of normal listeners with acoustic stimulation. Nevertheless, by using a method of amplitude normalization, Shannon showed a time course of recovery from adaptation in these subjects that was similar to that of normal subjects. He suggested that the primary source of forward masking is the central nervous system rather than the auditory periphery.

Similarities in psychophysical and peripheral physiological adaptation support the notion that the former is based on the latter. Nevertheless, the quantitative differences suggest that other factors are involved. Central nervous system effects are evident in physiological recordings. These central effects likely contribute to results of perceptual experiments on adaptation. Quantitative modeling of specific central adaptation mechanisms has yet to be accomplished.

REFERENCES

1. B. Scharf, in J. V. Tobias and E. D. Schubert (Ed.), *Loudness Adaptation in Hearing Research and Theory*, Academic, New York, 1983.

2. J. D. Hood, "Studies in Auditory Fatigue and Adaptation," *Acta OtoLaryngol.*, Suppl. 92, 1950.
3. T. E. Stokinger, W. A. Cooper, and W. A. Meissner, "Influence of Binaural Interaction on the Measurement of Perstimulatory Loudness Adaptation," *J. Acoust. Soc. Am.*, Vol. 51, 1972, pp. 602–607.
4. D. E. Morgan and D. D. Dirks, "Suprathreshold Loudness Adaptation," *J. Acoust. Soc. Am.*, Vol. 53, 1973, pp. 1560–1565.
5. S. A. Feaster and E. M. Weiler, "The effects of Monaural Auditory Adaptation in the Speech Range," *Br. J. Audiol.*, Vol. 9, 1975, pp. 81–836.
6. G. Canevet, B. Schard, and M.-C. Botte, "Simple and Induced Loudness Adaptation," *Audiology*, Vol. 24, 1985, pp. 430–436.
7. I. K. Samoilova, "Masking of Short Tone Signals as a Function of the Time Interval Between Masked and Masking Sounds," *Biophysics*, Vol. 4, 1959, pp. 44–52.
8. S. P. Bacon and N. F. Viemeister, "The Temporal Course of Simultaneous Tone-on-Tone Masking," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 1231–1235.
9. D. M. Green, "Masking with Continuous and Pulsed Sinusoids," *J. Acoust. Soc. Am.*, Vol. 46, 1969, pp. 939–946.
10. S. P. Bacon and N. F. Viemeister, "Simultaneous Masking by Gated and Continuous Sinusoidal Maskers," *J. Acoust. Soc. Am.*, Vol. 78, 1985, pp. 1220–1230.
11. E. Zwicker, "Temporal Effects in Simultaneous Masking by White-Noise Bursts," *J. Acoust. Soc. Am.*, Vol. 37, 1965, pp. 653–663.
12. S. P. Bacon and M. A. Smith, "Spectral Intensive and Temporal Factors Influencing Overshoot," *Quart. J. Exper. Psychol.*, Vol. 43A, 1991, pp. 373–399.
13. S. P. Bacon, "Effect of Masker Level on Overshoot," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 698–702.
14. D. M. Green, M. J. McKey, and J. C. R. Lichlider, "Detection of a Pulsed Sinusoid in Noise as a Function of Frequency," *J. Acoust. Soc. Am.*, Vol. 31, 1959, pp. 1146–1452.
15. N. F. Viemeister, "Adaptation of Masking," in G. van den Brink and F. A. Bilsen, (Eds.), *Psychophysical, physiological, and Behavioral Studies in Hearing*, Delft University Press, Delft, The Netherlands, 1980, pp. 190–199.
16. D. McFadden and B. A. Wright, "Temporal Decline of Masking and Comodulation Detection Differences," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 711–724.
17. D. McFadden, "Comodulation Detection Differences Using Noiseband Signals," *J. Acoust. Soc. Am.*, Vol. 81, 1987, pp. 1519–1527.
18. J. J. Zwislocki, E. Pirodda, and H. Rubin, "Post-Stimulatory Effects at Threshold," *J. Acoust. Soc. Am.*, Vol. 31, 1952, p. 9.
19. H. Duifhuis, "Consequences of Peripheral Frequency Selectivity for Nonsimultaneous Masking," *J. Acoust. Soc. Am.*, Vol. 54, 1973, pp. 1471–1488.
20. C. P. Widin and N. F. Viemeister, "Intensive and Temporal Effects in Pure-Tone Forward Masking," *J. Acoust. Soc. Am.*, Vol. 66, 1979, pp. 388–395.
21. M. J. Penner, E. Cudahy, and G. W. Jenkins, "The Effect of Masker Duration on Forward and Backward Masking," *Percept. Psychophys.*, Vol. 15, 1974, pp. 405–410.
22. B. C. J. Moore, "Psychophysical Tuning Curves Measured in Simultaneous and Forward Masking," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 524–532.
23. S. P. Bacon and W. Jesteadt, "Effects of Pure-Tone Forward Masker Duration on Psychophysical Measures of Frequency Selectivity, *J. Acoust. Soc. Am.*, Vol. 82, 1987, pp. 1925–1932.
24. I. J. Russell and P. M. Sellick, "Intracellular Studies of Hair Cells in the Mammalian Cochlea," *J. Physiol.*, Vol. 284, 1978, pp. 261–280.
25. N. Y. S. Kiang, T. Watanabe, E. C. Thomas, and L. F. Clark, "Discharge Patterns of Single Fibers in the Cat's Auditory Nerve," Research Monograph No. 35, MIT Press, Cambridge, MA, 1965.
26. W. S. Rhode and P. H. Smith, "Characteristics of Tone-Pip Response Patterns in Relationship to Spontaneous Rate in Cat Auditory Nerve Fibers," *Hear. Res.*, Vol. 18, 1985, pp. 159–168.
27. R. L. Smith and M. L. Brachman, "Adaptation in Auditory-Nerve Fibers: A Revised Model," *Biol. Cybernet.*, Vol. 44, 1982, pp. 107–120.
28. L. A. Westerman and R. L. Smith, "Rapid and Short-Term Adaptation in Auditory Nerve Responses," *Hear. Res.*, Vol. 15, 1984, pp. 249–260.
29. T. C. Chimento and C. E. Schreiner, "Adaptation and Recovery from Adaptation in Single Fiber Responses of the Cat Auditory Nerve," *J. Acoust. Soc. Am.*, Vol. 90, 1991, pp. 263–273.
30. R. L. Smith and M. L. Brackman, "Dynamic Response of Single Auditory Nerve Fibers: Some Effects of Intensity and Time," in G. van den Brink and F. A. Bilsen (Eds.), *Psychophysical, Physiological, and Behavioral Studies in Hearing*, Delft University Press, Delft, The Netherlands, 1980.
31. R. L. Smith and J. J. Zwislocki, "Short-Term Adaptation and Incremental Responses of Single Auditory Nerve Fibers," *Biol. Cybernet.*, Vol. 17, 1975, pp. 169–182.
32. D. M. Harris and P. Dallos, "Forward-Masking of Auditory Nerve Fiber Responses," *J. Neurophysol.*, Vol. 42, 1979, pp. 1083–1107.
33. R. L. Smith, "Short-Term Adaptation in Single Auditory-Nerve Fibers: Some Post-Stimulatory Effects," *J. Neurophysiol.*, Vol. 40, 1977, pp. 1098–1112.
34. P. J. Abbas, "Effects of Stimulus Frequency on Adaptation in Auditory-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 65, 1979, pp. 162–165.
35. M. C. Liberman, "Auditory-Nerve Responses from Cats Raised in a Low-Noise Chamber," *J. Acoust. Soc. Am.*, Vol. 63, 1978, pp. 442–455.
36. E. M. Relkin and J. R. Doucet, "Recovery from Prior Stimulation I: Relationship to Firing Rate of Primary

Auditory Neurons," *Hear. Res.*, Vol. 55, 1991, pp. 215–222.

37. N. Y. S. Kiang, E. Moxon, and A. Kahn, "The Relationship of Gross Potentials Recorded from the Cochlea to Single Unit Activity in the Auditory Nerve," in R. Ruben, C. Elberling, and G. Salomon (Eds.), *Electrocochleography*, University Park, Baltimore, MD, 1976, pp. 95–115.
38. A. R. Moller and P. J. Jannetta, "Neural Generation of the Auditory Brainstem Response," in J. T. Jacobson (Ed.), *The Auditory Brainstem Response*, College Hill, San Diego, 1985, pp. 13–31.
39. T. C. Chimento and C. E. Schreiner, "Time Course of Adaptation and Recovery from Adaptation in the Cat Auditory Nerve Neurophonic," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 857–864.
40. M. Relkin and J. R. Doucet, "Peristimulus Compound Action Potential Recorded from the Auditory Nerve with a Suction Electrode," *Abst. 12th ARO Midwinter Meeting*, 1990, p. 119.
41. P. J. Abbas, "Recovery from Long-Term and Short-Term Adaptation of the Whole-Nerve Action Potential," *J. Acoust. Soc. Am.*, Vol. 75, 1984, pp. 1541–1547.
42. P. J. Abbas and M. P. Gorga, "AP Responses in Forward-Masking Paradigms and Their Relationship to Responses of Auditory-Nerve Fibers," *J. Acoust. Soc. Am.*, Vol. 69, 1981, pp. 492–499.
43. E. M. Relkin and R. L. Smith, "Forward Masking of the Compound Action Potential: Thresholds for the detection of the N_1 Peak," *Hear. Res.*, Vol. 53, 1991, pp. 133–140.
44. P. Dallos and M. A. Cheatham, "Compound Action Potential (AP) Tuning Curves," *J. Acoust. Soc. Am.*, Vol. 59, 1976, pp. 591–597.
45. M. Gorga, J. McGee, E. Walsh, E. Javel, and G. Farley, "ABR Measurements in the Cat Using a Forward Masking Paradigm," *J. Acoust. Soc. Am.*, Vol. 73, 1983, pp. 256–261.
46. T. Furukawa and S. Matsuura, "Adaptive Rundown of Excitatory Post-Synaptic Potentials at Synapses between Hair Cells and Eight Nerve Fibres in the Goldfish," *J. Physiol. (London)*, Vol. 276, 1978, pp. 193–209.
47. E. Javel, Y. C. Tong, and R. K. Shepherd, and G. M. Clark, "Responses of Cat Auditory Nerve Fibers to Biphasic Current Pulses," *Ann. Oto. Rhinol. Laryngol.*, Vol. 96, Suppl. 128, 1987, pp. 26–30.
48. C. van den Honert and P. H. Stypulkowski, "Temporal Patterns of Single Auditory Nerve Fibers Elicited by Periodic Electrical Stimuli," *Hear. Res.*, Vol. 29, 1987, pp. 207–222.
49. M. P. Schroeder and J. L. Hall, "Model for Mechanical to Neural Transduction in the Auditory Receptor," *J. Acoust. Soc. Am.*, Vol. 55, 1974, pp. 1055–1060.
50. H. A. Schwid and C. D. Geisler, "Multiple Reservoir Model of Neurotransmitter Release by a Cochlear Inner Hair Cell," *J. Acoust. Soc. Am.*, Vol. 72, 1982, pp. 1435–1440.
51. R. Meddis, "Simulation of Mechanical to Neural Transduction in the Auditory Receptor," *J. Acoust. Soc. Am.*, Vol. 79, No. 3, 1986, pp. 702–711.
52. R. Meddis, "Simulation of Auditory-Neural Transduction: Further Studies," *J. Acoust. Soc. Am.*, Vol. 83, No. 3, 1988, pp. 1056–1063.
53. J. J. Eggermont, "Peripheral Auditory Adaptation and Fatigue: A Model Oriented Review," *Hear. Res.*, Vol. 18, 1985, pp. 57–71.
54. E. M. Relkin and C. W. Turner, "A Reexamination of Forward Masking in the Auditory Nerve," *J. Acoust. Soc. Am.*, Vol. 84, 1988, pp. 584–591.
55. W. M. Seibert, "Stimulus Transformations in the Peripheral Auditory System," in P. A. Kolers and M. Eden (Eds.), *Recognizing Patterns*, MIT Press, MA, 1968.
56. E. D. Young and M. B. Sachs, "Recovery of Detection Probability Following Sound Exposure: Comparison of Physiology and Psychophysics," *J. Acoust. Soc. Am.*, Vol. 54, 1973, pp. 1544–1553.
57. D. McFadden and C. A. Champlin, "Reductions in Overshoot During Aspirin Use," *J. Acoust. Soc. Am.*, Vol. 87, 1990, pp. 2634–2642.
58. C. A. Champlin and D. McFadden, "Reductions in Overshoot Following Intense Sound Exposures," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 2005–2011.
59. F. G. Zeng, C. W. Turner, and E. M. Relkin, "Recovery from Prior Stimulation II: Effects Upon Intensity Discrimination," *Hear. Res.*, Vol. 55, 1991, pp. 223–230.
60. D. McFadden, "Spectral Differences in the Ability of Temporal Gaps to Reset the Mechanisms Underlying Overshoot," *J. Acoust. Soc. Am.*, Vol. 85, 1989, pp. 254–261.
61. R. V. Shannon, "Forward Masking in Patients with Cochlear Implants," *J. Acoust. Soc. Am.*, Vol. 88, 1990, pp. 741–744.

123

HEARING THRESHOLDS

WILLIAM A. YOST AND MEAD C. KILLION

1 INTRODUCTION

Thresholds of hearing define the lowest sound levels (absolute hearing thresholds) that a listener can detect as well as the highest (upper limits of audibility) sound levels that a listener can tolerate. These thresholds are used to describe hearing sensitivity and the dynamic range of hearing for both normal and hearing-impaired listeners. This chapter reviews the presentation and calibration procedures for sound field measurements and earphone measurements used to obtain absolute hearing thresholds to pure tones. Absolute thresholds obtained in a sound field are used to estimate minimal audible field (MAF) thresholds, while those absolute thresholds obtained when the sounds are presented over earphones are used to estimate minimum audible pressure (MAP) thresholds. The chapter compares the relationship between (absolute) MAF and MAP thresholds and provides a brief discussion of the psychophysical procedures used to obtain these thresholds. The chapter ends with a brief description of the thresholds of feeling and discomfort that described the upper limits of audibility and allows for the calculation of the dynamic range of hearing.

2 THRESHOLDS

Absolute hearing thresholds are the lowest sound pressure levels (SPLs) required for listeners to detect sound. The stimuli most frequently employed to obtain hearing thresholds and therefore to measure hearing sensitivity are pure tones in the frequency range from 500 through 4000 Hz. Although complex stimuli such as frequency-modulated tones and narrow-band noises can also be used to measure hearing thresholds, this chapter will describe procedures for the use of pure tones. For audiological evaluations testing is usually performed with a pure-tone audiometer, an instrument for measuring hearing sensitivity that provides pure tones of selected frequencies at calibrated SPLs. The results of such testing are recorded on an audiogram, a graph showing hearing (threshold) level as a function of frequency. *Hearing* (*threshold*) *level* (HL) is the number of decibels that the listener's threshold of hearing lies above the zero reference level of the audiometer for that frequency. Thus, *0 dB HL* (*hearing level*) represents this zero reference level, which is based on the average hearing level of a large number of young adults considered to be otologically normal. The SPLs corresponding to 0 dB HL have standardized both nationally and internationally.[1,2]

In most cases, the listener's ability to hear everyday sounds in normal rooms or outdoors is ultimately of primary interest, but that ability is difficult to test directly for obvious reasons. Measurements of the thresholds of hearing that employ pure tones allow for the easy and accurate reporting and interpretation of the status of an individual's hearing, frequency by frequency, in each ear. Two ways of presenting sounds are typically used: Sounds can be presented over loudspeakers with the listener in the sound field or over earphones fitted to the listener (earphones are the more frequent method used to obtained thresholds of hearing). For both ways of presenting sounds, the SPLs are calibrated so as to produce 0 dB HL thresholds for the average normal listener. It is, however, difficult to determine directly the ability, willingness, and response tendencies of the individual in the test situation. *Psychophysical procedures* prescribe

---

Encyclopedia of Acoustics, edited by Malcolm J. Crocker
ISBN 0-471-80465-7 © 1997 John Wiley & Sons, Inc.

the methods of sound presentation and threshold calculation used to estimate thresholds. These psychophysical procedures help ensure that the obtained thresholds estimate the lowest SPL a listener can detect at each test frequency.

Given that hearing levels are referenced to "normal hearing," any two thresholds determined on a given listener might be expected to agree, regardless of the method used to estimate the threshold or how the tones were presented to the listener. In practice, the accuracy and reliability (correctness and repeatability) of a given threshold depend on several factors:

1. Accuracy of calibration of the sound source, which is typically performed with the listener absent
2. Idiosyncratic effect of the individual listener's ear (and head in the case of sound field testing) on the stimulus delivered to the listener's eardrum by the sound source.
3. Accuracy of the zero reference SPL used for the calibration of the sound source
4. Level of background noise entering the listener's ear canal
5. Listener's willingness and ability to cooperate in the task of determining the level of sound for threshold
6. Method of stimulus presentation
7. Response procedure (type of psychophysical procedure)
8. Number of responses averaged to obtain a single threshold estimate

The eight factors listed above fall into two categories: (1) sound presentation and calibration and (2) threshold calculation method. The first yields an estimate of the SPL entering the auditory system under the measurement conditions, and the second yields an estimate of the lowest sound pressure level at which "hearing" occurs, the threshold. Discussions of "threshold" sometimes fail to make clear the distinction between *absolute hearing threshold*, measured in units of the minimum SPL required for audibility, and *relative hearing threshold level* (HL), measured in units of the number of decibels above the average normal hearing threshold SPL at that frequency. Figure 1 summarizes the various thresholds discussed in this chapter: minimum-audibility thresholds (MAF and MAP thresholds) and thresholds for discomfort and feeling, which estimate the upper limits of audibility.

3 SOUND FIELD MEASUREMENTS

Minimum-audible-field thresholds are SPLs for pure tones at absolute threshold, measured in a free field at the position of the listener's head but in the absence of the listener, using plane, progressive sound waves. The MAF thresholds are usually determined for listeners facing the source (0° incidence), listening with both ears (binaurally), at 1 m from the sound source, and are based on progressive plane-wave assumptions. If the loudspeaker is approximately 1 m from the measurement microphone [the standard method for calibrating microphones can be found in ANSI S1.10-1966(R1986)[9]], then the difference in diffraction around the head between a progressive plane wave and a progressive spherical wave will be negligible.

The acoustics of the room (reflections and standing waves) in which the measurements will be made can affect the sound as it travels from the loudspeaker to the listener. An anechoic room may reduce standing waves, especially at high frequencies, although most anechoic rooms do not eliminate reflections at very low frequencies (below 200 Hz). If other sound field conditions are used, they should be specified. For example, if thresholds are measured in a *diffuse sound field*, then they should be identified as *MAF* (*diffuse-field*) thresholds. In a diffuse sound field sound arrives at the listener from *all* angles (including above and below) with equal probability (see ISO/R226-1985[2]).

Under plane-wave conditions, the pressure at one ear depends on the angle of the sound source relative to the ear of the listener. The data in Fig. 2 show the MAF[10] thresholds obtained for frequencies ranging from 300 to 15,000 Hz and at azimuth (horizontal plane) angles circling the head (0° represents the condition in which the speaker is directly in front of the listener). Binaural MAF thresholds obtained when both ears are used may be slightly lower (3 dB or less) than monaural MAF thresholds obtained when the listener uses his or her more sensitive ear. Thus, both the type of sound field and the monaural or binaural listening condition influence the threshold estimates when free-field MAF thresholds are obtained or reported.

To achieve a monaural listening condition, one ear of a listener must be covered with a hearing protector. The ANSI standards S3.19-1974(R1979)[11] and S12.6-1984(R1990)[12] specify the measurement methods for determining the attenuation characteristics of hearing protectors. A combination of an in-the-ear hearing protector and an over-the-ear protector may be used to achieve the greatest amount of attenuation. A masking noise can also be presented over an earphone to one ear in order to mask sounds arriving at that ear. The ANSI

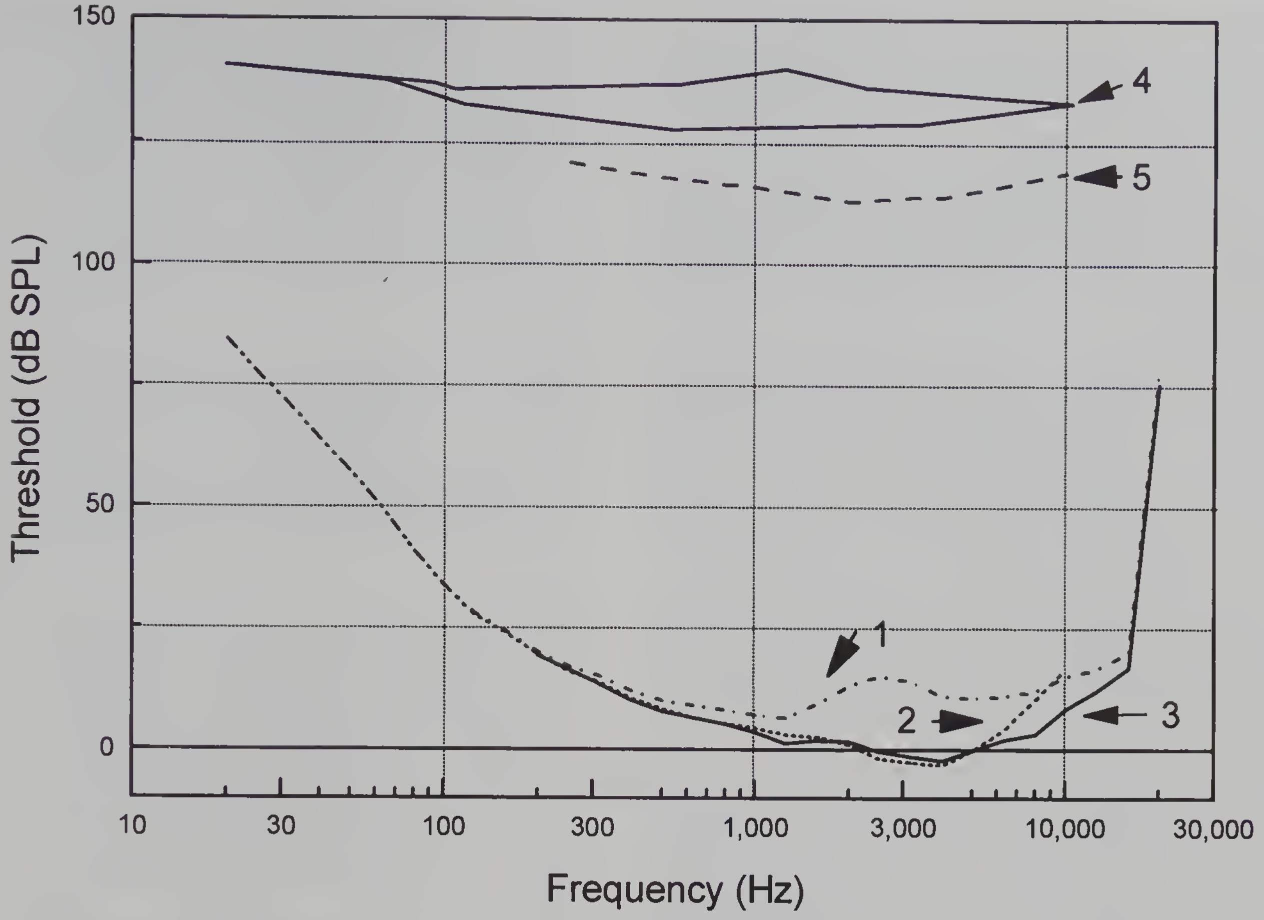


Fig. 1 Thresholds of hearing (minimum-audibility thresholds) expressed in decibels SPL shown as a function of the test frequency for three conditions: curve 1, binaural MAF thresholds,[3] obtained for 0° azimuth; curve 2, monaural MAP thresholds[3]; curve 3, thresholds obtained in a diffuse sound field MAF(Diffuse Field). Curves 4 and 5 represent thresholds associated with the upper limits of audibility. The thresholds shown as curve 4 represents the range of thresholds when listeners report the level that led to a feeling of "tickle," "pain," or "feeling."[4–7] Thresholds of discomfort are shown as curve 5.[6–8]

standard S3.6-1989[1] describes the sound levels of noise that are effective for masking the sounds at one ear while allowing sounds to be detected at the other ear.

Because the angle of incidence of the sound is so important, small head movements can often cause large changes in the sound pressure presented to the eardrum in a MAF procedure. This is particularly true in the 6–8-kHz region for free-field sounds of 0° incidence (approximately plane wave), where a null-in-eardrum SPL of 20–30 dB can occur due to an antiresonance excited in the concha.[13]

The MAF thresholds from two conditions are shown in Fig. 1: (1) binaural MAF thresholds obtained for 0° azimuth and (2) binaural MAF (diffuse-field) thresholds. These curves were obtained by combining the data for MAP thresholds (discussed below) with data relating eardrum SPL with sound field and SPLs.[3,14,15] Column 1 of Table 1 provides the International Standardization Organization (ISO) standardized MAF thresholds,[2] with low-frequency corrected values[3,14] shown in column 2. Additional studies[16] indicate that the ISO standard thresholds may underestimate 0-HL levels by 1–4 dB above 500–1000 Hz. On the other hand some recent studies tend to support the uncorrected ISO values at low frequencies.[16]

4 EARPHONE MEASUREMENTS

Although thresholds obtained from earphone presentations are often referred to as MAP thresholds, for the sake of clarity MAP thresholds will be restricted to the following definition,[10] "the [sound] pressure amplitude at the observer's eardrum." The MAP curve in Fig. 1 is an estimate of the minimum audible sound pressure at the eardrum required for threshold.[3,15]

Although the minimum audible pressure at the eardrum is independent of the type of (airborne) sound source,[3,17] an audiometric earphone is normally not calibrated in terms of the eardrum pressure it produces but

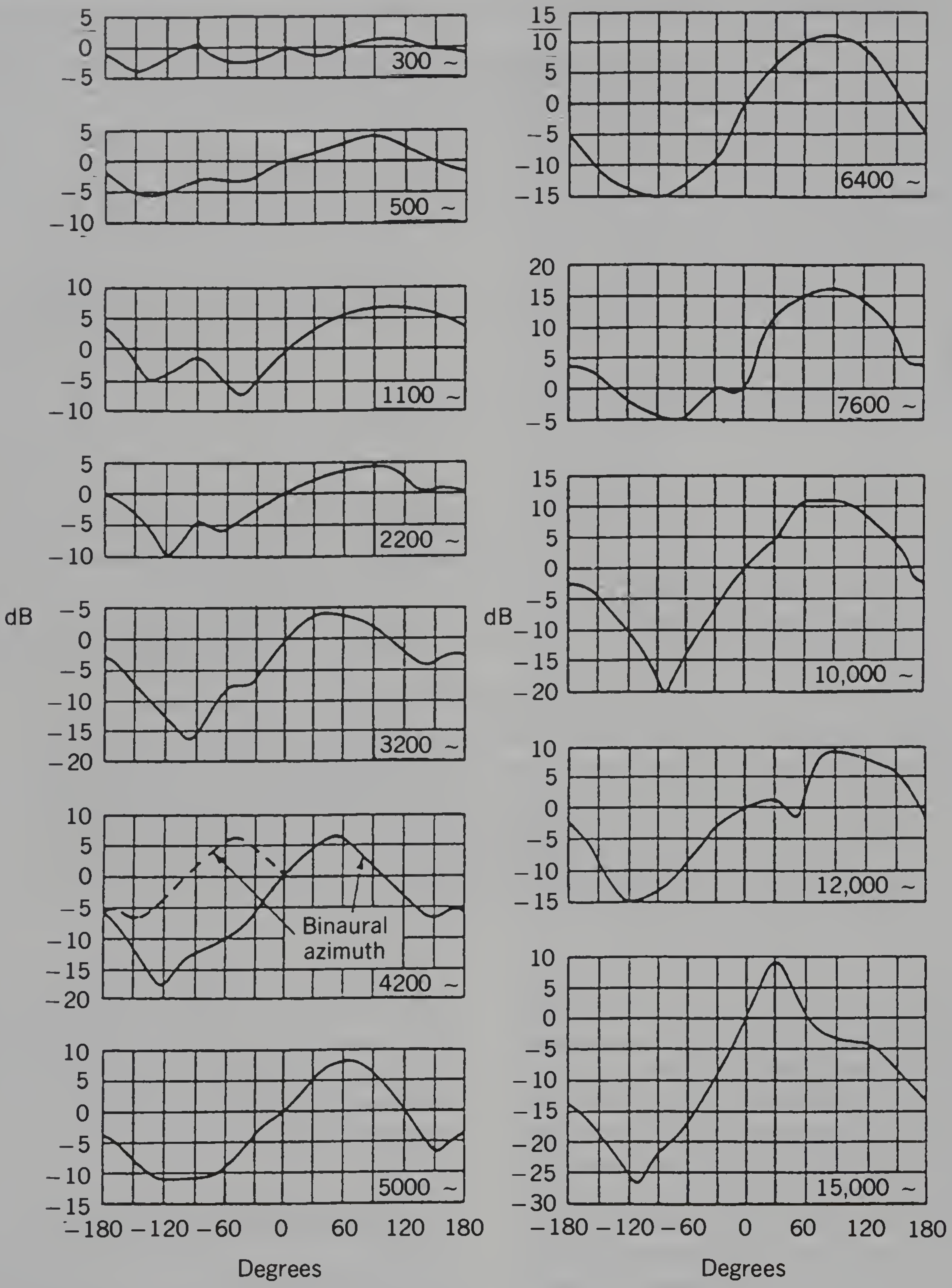


Fig. 2 Minimum audible field thresholds^2 for one ear are shown as a function of the azimuth angle between the listener and the loudspeaker (0° is directly in front of listener, 180° directly behind). Each panel represents a different test frequency. The MAF thresholds for each azimuth angle are referenced to that measured at 0°.

in terms of the *coupler* pressure it produces when the *average-normal-threshold volage* is applied to the earphone terminals. That is, at the eardrum, the SPL is estimated from the sound level in a test coupler attached to the earphone. Such couplers are typically designed to roughly approximate the average acoustic properties of the ear of a normal hearing listeners. The zero reference SPLs for a particular earphone–coupler combination are called the *reference equivalent threshold sound pressure levels* (RETSPLs) and are given in the earphone calibration tables of audiometry standards, such as shown in Table 2. Table 2 shows the RETSPLs[1] given in ANSI

TABLE 1 Minimum Audible Field Thresholds

Frequency (Hz)	MAF Thresholds (dB SPL)	Corrected
100	25.1	33
125	20.7	30
160	16.8	24
200	13.8	18.5
250	11.2	17
315	8.9	12
400	7.2	8
500	6.0	
630	5.0	
800	4.4	
1000	4.2	
1250	3.7	
1600	2.6	
2000	1.0	
2500	−1.2	
3150	−3.6	
4000	−3.9	
5000	−1.1	
6300	6.6	
8000	15.3	
10000	16.4	
12500	11.6	

Note: Minimum audible field (MAF) thresholds are expressed in decibels SPL as specified in the International Standardization Organization's acoustic standard ISO/R226-1985 (Ref. 2). The conditions that must exist for these thresholds to apply are (1) the source is at 0° azimuth (frontal incident), (2) the sound field is a progressive plane wave, (3) listening is binaural, and (4) the listeners are otologically normal people in the age from 18 to 30 years inclusive. Corrected values are shown for certain low frequencies (from Ref. 3).

S3.6-1984. Note that each earphone–coupler combination generates a different set of values for RETSPL, only one of which is equal to the estimate of MAP (column 5). That earphone–coupler combination, the Zwislocki coupler occluded-ear simulator (described in more detail below), is designed to provide the best simulation of the average ear and eardrum.

4.1 Supra-aural Earphones

Supra-aural earphones are calibrated in a simple "6-cc" coupler such as the NBS-9A coupler [ANSI 3.7-1973(R1986)[18] specifies the dimensions for this coupler]. The coupler has a volume of 6 cc, an approximation to the average volume of the normal human outer ear between the earcap of the earphone and the eardrum. During calibration, the earcap of the earphone rests on the upper edge of the coupler and is held in place by a force of between 400*g* and 500*g*. A calibrated microphone[9] measures the SPL.

The NBS-9A coupler only roughly simulates the acoustic impedance of the average male outer ear covered with a supra-aural earphone. Based on the real-ear, probe–microphone data,[13] the actual eardrum pressure may range between 1 and 9 dB greater than the coupler pressure over the frequency range of approximately 1000–3000 Hz.[1] At frequencies below 400 Hz, on the other hand, the actual eardrum pressure produced by a supra-aural earphone will usually be substantially less than the measured coupler pressure due to leakage of sound from around the earphone fitted to the ear. Thus, care must be used in fitting the earphone to the outer ear. A headband with a spring to hold the earphone cushion against the ear should be used. Jewelry, eye-glasses, and hair from the head should not lie between the earphone cushion and the ear. The earphone must be fitted with a proper cushion (see ANSI S3.6-1989[1]) to help ensure a good fit to the ear. When measurements from both ears are to be made, both earphones are calibrated and the differences between earphones is kept as small as possible (the two phones should differ by less than 2.5 dB for each frequency tested).

Supra-aural earphone thresholds obtained at frequencies above 8000 Hz may also vary considerably (10 dB or more) from listener to listener due to standing waves within the outer ear. That is, because the length and volume of the outer ear are of approximately the same size as the quarter wavelength of these high-frequency tones, standing waves may be produced that lead to nonuniform distribution of sound pressure within the outer ear. A number of techniques have been used to obtain MAP thresholds above 8000 Hz and to calibrate the expected sound pressure within the outer ear. However, none of these techniques has been standardized and large differences in RETSPL values (on the order of 30 dB) exist among the thresholds obtained across studies.[20]

The ambient background noise level at the location of the listener must be no higher than the values described in ANSI S3.1-1977(R1991).[21] Background noise is likely to be especially deleterious for estimating thresholds below 1000 Hz using supra-aural earphones. Thus, spura-aural earphone thresholds should be measured in a sound-proof room meeting the acoustical requirements of ANSI 3.1-1977(R1991).[21] In addition to ambient background noise, thresholds obtained with supra-aural earphones at low frequencies (500 Hz and below) are often affected by physiological noise in the outer ear. This noise is caused by the sound of blood circulating in the arteries and veins of the head and neck and by sounds associated with respiration and body movements.

4.2 Insert Earphones

Insert earphones sealed directly into the ear canal with deeply inserted foam eartips of the type used for hearing protection (but including a sound tube through the

TABLE 2 Reference Equivalent Threshold Sound Pressure Level (RETSPL) Thresholds for Different Earphone–Coupler Combinations

	RETSPL Thresholds (dB SPL)					
Frequency (Hz)	Western Electric 705A[a] (1)	Telephonics TDH 39, 39P[a] (2)	Telephonics TDH 49, 49P; TDH 50, 50P[a] (3)	Telex 147A[a] (4)	Insert[b] (5)	Etymontic ER 3A[c] (6)
125	45.5	45	47.5	47	30	27.5
250	24.5	25.5	26.5	27.5	19	15.5
500	11	11.5	13.5	13	12	8.5
750	7.5	8	8.5	8.5		
1000	6.5	7	7.5	6.5	9	3.5
1500	6.5	6.5	7.5	5		
2000	8.5	9	11	8	15	6.5
3000	7.5	10	9.5	7.5	15.5	5.5
4000	9	9.5	10.5	9	13	1.5
6000	8	15.5	13.5	17.5	13	−1.5
8000	9.5	13	13	17.5	14	−4.0

[a] Supra-aural earphone.

[b] Insert, occluded ear simulator.

[c] Insert, HA-1 coupler.

Note: From ANSI, S3.6-1989 (Ref. 1); columns 1–4 are from Table 6, column 5 from Table G.1, and column 6 from Table G.2. The listeners are otologically normal, the supra-aural earphones have been fitted with the proper cushion (described in Ref. 1), and the earphones have been calibrated via an acoustic coupler meeting the requirements of ANSI S3.7-1973(R1986) (Ref. 18) or 3.25-1979 (R1986) (Ref. 19).

center of the eartip) reduce the problem of the probable masking effect of physiological noise in the outer ear that occurs when supra-aural earphones are used. Such a deeply sealed eartip can also provide substantially more attenuation of ambient noise than that provided by supra-aural earphones and cushions.[22] Insert earphones are usually better at controlling leakage and acoustic cross-talk between the phones at each ear than are supra-aural earphones.[22]

Calibration of insert earphones is performed with the same couplers used to acoustically calilbrate hearing aids, using either a *2-cc coupler* like the HA-1 coupler or an *occluded-ear simulator* such as the Zwislocki coupler (see ANSI S3.25-1989[19]). The 2-cc coupler was designed to fit over a standard 1-in. (2.54-cm) microphone and to have dimensions that minimize troublesome standing waves at high frequencies. The 2-cc coupler roughly approximates the volume of air between the tip of an insert earphone and the eardrum, and it provides the equivalent volume of normal eardrum compliance at low frequencies. An *occluded-ear simulator* provides a closer approximation than the 2-cc coupler to the acoustic impedance of the average adult occluded-ear canal and eardum. In addition, the occluded-ear simulator provides an approximation to the standing-wave field that exists in the average adult outer ear. Because the insert earphone presents a high-impedance source to the ear, a constant-decibel relationship exists at any frequency between the SPL in a 2-cc coupler and that in an occluded-ear simulator. Thus, either calibration may be used with equivalent results.[23] The 2-cc coupler RETSPL values for the ER-3A insert earphone have been standardized, and these calibration values are given in column 6 of Table 2.

5 RELATIONSHIP BETWEEN (ABSOLUTE) MAF AND MAP THRESHOLDS

5.1 Stimulus Conditions

In all procedures absolute thresholds are expressed as decibels of sound pressure level (dB SPL): x dB SPL $= 20 \log_{10}(y/20\ \mu\text{Pa})$, where y is the pressure of the test signal measured in units of micropascals. The referent value of 20 μPa is approximately the SPL required for the average young adult to just detect a tone with test frequencies in the range of 1000–3000 Hz.

For most measurements the tonal signals used to obtain thresholds are pulsed sinuosids of a single frequency (within ±3% of the nominal frequency), produced with low total harmonic distortion (usually less than 3%), having a duration of at least 400 ms, and turned on and off with rise–decay times of at least 20 ms in order to reduce audible clicks and to further restrict spectral energy to the region of the tonal frequency. The frequencies of the tones are usually within the range of 20–20,000 Hz and are often 125 Hz and the six higher octaves of 125 Hz. Absolute thresholds remain approximately constant for durations longer than 300 ms. The exact relationship between thresholds 'and signal duration is frequency dependent.[24] Absolute thresholds are also known to increase with the age of the listener, and women often have slightly lower thresholds than men, especially for an older population.[25] There is also evidence that for test frequencies above 2000 Hz the right ear of many listeners is slightly more sensitive than the left ear.[25] In addition, the otologic condition of the listener will obviously affect the threshold estimated from a hearing-impaired listener.

5.2 MAF Thresholds Are Lower Than MAP Thresholds

The MAF and MAP thresholds do not agree when they are actually measured (see Fig. 1). For instance, head diffraction and external-ear resonances combine to cause some 15–20 dB increase in eardrum pressure over that in a frontally incident sound field at 2700 Hz. Since the sound field pressure is calibrated with the listener absent, this increase causes an exactly comparable decrease in sound field pressure required for threshold, that is, in the MAF data. The MAF thresholds are thus always lower than MAP thresholds by an amount that reflects the gain of the external ear at each frequency plus the binaural advantage (of 3 dB or less) when binaural MAF is compared to monaural MAP.

By examining the difference between the MAF and MAP thresholds, one might conclude that a "missing 6 dB[3,17]" existed between sound field (MAF) and earphone (MAP) thresholds, even when both had been converted to sound pressure at the eardrum (see the difference between curves 1 and 2 in Fig. 1). As Fig. 3 indicates, there are experimental difficulties that lead to that erroneous conclusion.[17] Figure 3 (solid curve) shows the MAP threshold estimate given in ANSI S3.6-1989,[1] based on independent estimates of MAP from free-field MAF data corrected for head diffraction and external-ear resonances and from earphone RETSPL data corrected for real ear–coupler differences.[3] Given the

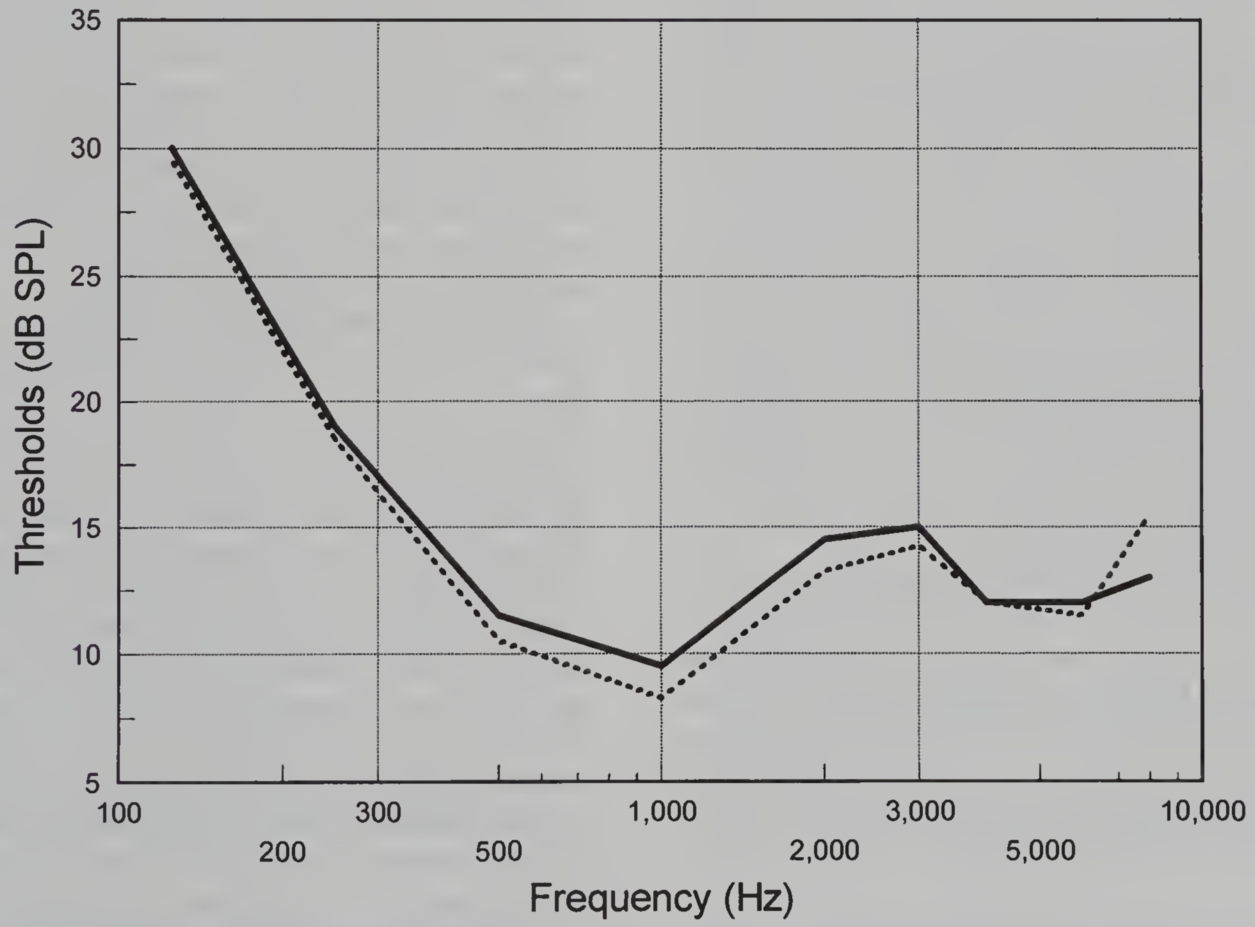


Fig. 3 Minimum audible pressure, RETSPL in decibels SPL as a function of test frequency. The solid curve shows corrected MAP thresholds from ANSI 3.6-1989,[10] and the dashed curve represents thresholds obtained with the ER-3A insert earphone converted to the equivalent occluded-ear simulator (Zwislocki-coupler) pressures.[3] The small difference between these two curves suggests there is no missing 6 dB.[3]

fixed relationship between 2-cc coupler sound pressure and occluded-ear coupler sound pressure at each frequency and the good agreement between sound pressure measured in real ears and in the occluded-ear coupler,[23] a relatively direct estimate of MAP can be made from insert earphone thresholds. Figure 3 (dashed curve) also shows MAP at the eardrum averaged from four studies in which 2-cc coupler RETSPL data for the ER-3A insert earphone were reported and subsequently converted to equivalent occluded-ear coupler pressure.[3] Thus, the MAP estimate shown as a solid curve in Fig. 3 was based on MAF and supra-aural earphone RETSPL data, while the dashed curve in Fig. 3 was based on insert earphone RETSPL data. The greatest difference between the two curves is 2.5 dB, indicating that the difference zero reference SPLs required for calibrating different sound sources are essentially the same. Any differences that are measured reflect only differences in the calibration procedure.[3] That is, there is no missing 6 dB.

The shape of the MAP and MAF absolute threshold functions are approximately the same, showing a loss of sensitivity below approximately 1000 Hz and above approximately 4000 Hz. The loss of sensitivity at low frequencies below 1000 Hz declines at approximately 6 dB/octave and at approximately 24 dB/octave above 4000 Hz (the ability to estimate the slope of the high-frequency side of the thresholds of hearing function is much poorer than estimating the low-frequency slope due to the difficulty in measuring thresholds above 8000 Hz, as discussed previously). A threshold of 0 dB SPL represents a pressure of 20 μPa, an extremely small pressure. By making some assumptions about the acoustic energy present in the Brownian motion of air molecules, it can be shown that a sound presented at 0 dB SPL is only 20–30 dB more intense than that being produced by Brownian motion.[26] A number of factors governing the acoustic transfer function of the outer and middle ears have been suggested as explanations for the shape of the thresholds of hearing functions.[27]

The data shown in Fig. 1 are based on test frequencies that are widely spaced. When thresholds of hearing are measured over a small range with test frequencies close together, the thresholds vary from test frequency to test frequency in a pattern that is unique to the individual being tested. This *microstructure* to the audiogram can show threshold changes of 5–10 dB that occur when the test frequency has changed by just a few hertz.[28,29]

6 PSYCHOPHYSICAL PROCEDURES

A number of psychophysical procedures and variations of these procedures have been used to obtain MAF or MAP thresholds. In general, sounds are presented to listeners at different SPLs, and listeners make judgments concerning the presence or absence of the sound. Each psychophysical procedure specifies a precise method for presenting the tones and obtaining a threshold from the functional relationship between the listener's responses and the SPLs presented to the listeners. Hearing thresholds obtained using different psychophysical procedures for the same sound stimulus may differ by 5 dB or more.[30]

Thresholds obtained from the various psychophysical procedures often depend both on the integrity of the auditory system and on other variables that affect how a listener responds. The variables that are not associated with auditory processing but may influence a listener's response are referred to as *response bias variables*.[30] For instance, the instructions given to a listener may make the listener liberal or conservative in the use of the response alternatives. Since thresholds depend on how the listener responds, response bias variables may affect the obtained threshold. *Forced-choice methods* are often the preferred method for reducing the effect of response bias variables on the obtained thresholds. In some forced-choice methods the tonal signal is presented half the time and no tonal signal is presented half the time, and the listener is to indicate whether the tonal signal or no tonal signal was presented.[30] Other methods that yield thresholds that may depend on respose bias but are often used because of their efficiency include the *method of limits* and the *method of adjustment*. In the method of limits, the tester decreases and then increases the level of the test stimulus and on each presentation asks the listener if he or she detected the test tone. In the *method of adjustment* procedure the listener adjusts the level of the test tone until it is just barely detectable.

Quite often one requires an efficient procedure that yields similar thresholds when the same listener is tested more than once. The typical clinical threshold procedure uses a method of adjustment, which takes approximately 45 s per threshold and yields a standard deviation for a single threshold estimate of about 3.5 dB.[31] Several studies[31] suggests that versions of automatic adjustment procedures, such as the Bekesy procedure, provide a more efficient procedure: on the order of 30 s per threshold with a standard deviation for a single threshold estimate of approximately 2 dB. The forced-choice procedures often produce the lowest thresholds and are intrinscially free of response bias, but they are the least efficient procedure, requiring on the order of 3 min per threshold estimate with a 3-dB standard deviation.[31]

7 THRESHOLDS OF FEELING AND DISCOMFORT

Thresholds of feeling and discomfort are used to estimate the highest SPLs that the human ear can tolerate and are used to define the *upper limits of audibility*. Thresholds estimating the upper limit of audibility are often obtained by increasing the sound level of intense sounds until listeners report that they have experienced certain sensations or perceptions that indicate the auditory system may be jeopardized. In most procedures the SPL of the tone at each frequency is adjusted upward, in a type of method of limits procedure, until the listener indicates that the sound level has caused the required sensation. The sound level is not increased any further to prevent risk to the auditory system, and this final sound level forms an estimate of the upper limit of audibility for that run, and the procedure is repeated again. The average final sound level for a number of runs (usually four or more) forms a threshold for any one frequency. The same stimulus conditions that were recommended for MAP and MAF threshold measurements are usually used for estimating the upper limit of audibility, and special care is required to ensure that the tones are not distorted by the sound presentation apparatus at these high levels.

A number of sensations indicate that the sound level has reached the upper limit that the ear can tolerate. A listener can experience a "tickle," "prickling," "feeling," or "pain" sensation.[4–7] Any of these sensations can be used as the response definition estimating the upper lilmit of audibility. However, asking listeners to indicate when they "feel" the sound in the ear is the most common criterion, and the other sensations and thresholds appear to be highly correlated with the sensation of feeling and the threshold of feeling.[6] Curves 4 at the top of Fig. 1 display the sound levels that lead to thresholds of feeling, pain, and tickle. Considerable variability is to be expected when obtaining thresholds estimating the upper limit of audibility due to listeners' uncertainties in making determinations (response bias) of pain, feeling, tickle, and so on.[6]

Because of the risk of permanently damaging the auditory system, *thresholds of discomfort* (those SPLs leading to a report of "discomfort") are often used to estimate the upper limits of audibility. For many purposes, (e.g., the fitting of hearing aids), the definition of discomfort may include a context for the listening situation. For instance, the listener may be asked to indicate when the sound level has reached the point where he or she would no longer wear a hearing aid because the sound level was uncomfortable. The same procedure used for determining the thresholds of feeling is usually used to obtain the thresholds of discomfort. Curve 5 in Fig. 1 indicates the thresholds of discomfort.[5,7,8] Thresholds of discomfort are less variable than, and 10–30 dB below, the thresholds of feeling[5]

The decibel difference between thresholds representing the upper limits of audibility and the thresholds of hearing define the *dynamic range of hearing*, which may be as large as 140 dB for frequencies between 1000 and 3000 Hz. The frequency range over which thresholds and estimates of the upper limits of audibility can be measured determine the *frequency range of audibility*, which ranges from 20 to 20,000 Hz for most young, otologically normal listeners. For most situations and listeners (e.g., for a hearing-impaired listener) changes in the dynamic range in any frequency region are due to changes in the thresholds of hearing.

REFERENCES

1. ANSI S3.6-1989, "American National Standard Specifications for Audiometers."
2. ISO/R226-1985, "International Standard Organization Normal Equal-Loudness Contours for Pure Tones under Free-Field Listening Conditions."
3. M. C. Killion, "Revised Estimates of Minimum Audible Pressure: Where Is the 'Missing 6 dB'?" *J. Acoust. Soc. Am.*, Vol. 63, No. 5, 1978, pp. 1501–1508.
4. R. L. Wegel, "Physical Data and Physiology of Excitation of the Auditory Nerve," *Ann. Otol. Rhinol. Laryngol.*, Vol. 41, 1924, pp. 740–779.
5. G. Von Bekesy, "Uber die Herstellung und Messung langsamer sinusformiger Luftdruckschwankungen," *Ann. Physik.*, Vol. 26, 1936, pp. 514–566.
6. S. R. Silverman, "Tolerance for Pure Tones and Speech in Normal and Defective Hearing," *Ann. Otol. Rhinol. Laryngol.*, Vol. 56, 1948, pp. 658–682.
7. J. C. R. Licklider, "Basic Correlates of the Auditory Stimulus," in S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, Wiley, New York, 1966.
8. H. W. Davis and R. S. Silverman, *Hearing and Deafness*, Holt Rinehart and Winston, New York, 1968.
9. ANSI S1.10-1966 (R 1986), "American National Standard Method for the Calibration of Microphones."
10. L. J. Sivian and S. D. White, "On Minimum Audible Fields," *J. Acoust. Soc. Am.*, Vol. 4, 1933, p. 288.
11. ANSI S3.19-1974 (R 1979), "American National Standard Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs.[11]
12. ANSI S12.6-1984(R1990), "American National Standard Method for the Measurement of Real-Ear Attenuation of Hearing Protectors."
13. E. A. G. Shaw, "Earcanal Pressure Generated by Circumaural and Supra-aural Earphones," *J. Acoust. Soc. Am.*, Vol. 39, No. 3, 1966, pp. 471–479.
14. E. H. Berger, "Re-examination of the Low-Frequency (50–1000 Hz) Normal Threshold of Hearing in Free and Diffuse Sound Fields," *J. Acoust. Soc. Am.*, Vol. 70, No. 6, 1981, pp. 1635–1645.
15. M. C. Killion, E. H. Berger, and R. A. Nuss, "Diffuse Field Response of the Ear," *J. Acoust. Soc. Am.*, Vol. 81, 1975, p. 575.

16. K. Betke, "New Hearing Threshold Measurements for Pure Tones under Free-Field Listening Conditions," *J. Acoust. Soc. Am.*, Vol. 89, No. 5, 1991, pp. 2400–2403.
17. W. Rudmose, "The Case of the Missing 6 dB," *J. Acoust. Soc. Am.*, Vol. 71, No. 3, 1982, pp. 650–659.
18. ANSI S3.7-1973 (R 1986), "American National Standard Method for Coupler Calibration of Earphones."
19. ANSI S3.25-1979 (R 1989), "American National Standard for an Occluded Ear Simulator."
20. D. M. Green, G. Kidd, and K. N. Stevens, "High-Frequency Audiometric Assessment of a Young Adult Poplation," *J. Acoust. Soc. Am.*, Vol. 81, No. 2, 1987, pp. 485–494.
21. ANSI 3.1-1977 (R 1991), "American National Standard Criteria for Permissible Ambient Noise during Audiometric Testing."
22. E. H. Berger and M. C. Killion, "Comparison of the Noise Attenuation of Three Audimetric Earphones, with Additional Data on Masking Near Threshold," *J. Acoust. Soc. Am.*, Vol. 86, No. 4, 1989, pp. 1392–1403.
23. R. M. Sachs and M. C. Burkhard, "Earphone Pressure Response in Real Ears and Couplers," Report 20022, Knowles Electronics, Elk Grove, IL, 1972.
24. C. S. Watson and R. W. Gengel, "Signal Duration and Signal Frequency in Relation to Auditory Sensitivity," *J. Acoust. Soc. Am.*, Vol. 46, No. 4, 1969, pp. 989–997.
25. D. Y. Chung, K. Mason, R. P. Gannon, and G. N. Wilson, "The Ear Effect as a Function of Age and Hearing Loss," *J. Acoust. Soc. Am.*, Vol. 73, No. 4, 1983, pp. 1277–1282.
26. D. M. Green, *An Introduction to Hearing*, Erlbaum, New York, 1976.
27. P. Dallos, *The Auditory Periphery*, Academic, New York, 1973.
28. E. Elliott, "A Ripple Effect in the Audiogram," *Nature*, 1958.
29. M. Cohen, "Detection Threshold Microstructure and Its Effects on Temporal Integration Data," *J. Acoust. Soc. Am.*, Vol. 71, No. 2, 1982, pp. 405–409.
30. D. E. Robinson and C. S. Watson, "Psychophysical Methods in Modern Psychophysics," in J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory*, Vol. 2, Academic Press, New York, 1972.
31. A. Hesse, "Comparison of Several Psychophysical Procedures with Respect to Threshold Estimates, Reproducibility and Efficiency," *Acustica*, Vol. 54, 1986, pp. 263–286.

LIBRARY USE ONLY

DATE DUE

JAN 2 5 2004

GAYLORD

PRINTED IN U.S.A.

UNIVERSITY OF CA. RIVERSIDE LIBRARY
3 1210 01200 0319

ISBN 0-471-18006-
4-VOL. SET ISBN 0-471-80465-

P6-AZA-053